PERIODICAL ROOM
GENERAL LIBRARY
UNIV. OF MICH.

OCT 21 1929

The Journal of EDUCATIONAL PSYCHOLOGY

Devoted Primarily to the Scientific Study of Problems of Learning and Teaching.

BOARD OF EDITORS
HAROLD RUGG, *Chairman.*

J. CARLETON BELL — ARTHUR I. GATES — RUDOLF PINTNER — L. M. TERMAN
FRANK N. FREEMAN — V. A. C. HENMON — BEARDSLEY RUML — E. L. THORNDIKE

VOL. XX
No. 7

OCTOBER, 1929

$5.00 a year.
70c per copy.

CONTENTS

CHOICE OF ITEMS FOR A TEST ON THE BASIS OF DIFFICULTY. *Percival M. Symonds* 481

THE SUBJECT-MATTER PROGRESS OF THREE ACTIVITY SCHOOLS IN HAWAII, WITH A NOTE ON STATISTICAL TECHNIQUE. *Helen G. Pratt, Jack W. Dunlap, and Edward E. Cureton* . . 494

A COMPARISON OF CERTAIN PHASES OF MUSICAL ABILITY OF COLORED AND WHITE PUBLIC SCHOOL PUPILS. *C. T. Gray and C. W. Bingham* 501

THE RELATIONSHIP OF MOTOR ABILITY TO THE INTELLIGENCE OF KINDERGARTEN CHILDREN. *Oscar E. Hertzberg* 507

CLINICAL OBSERVATION ON EYE MOVEMENTS IN THE RATING OF READING ABILITY. *W. R. Miles and David Segel* 520

SOME SCIENTIFIC APPROACHES TOWARD VOCATIONAL GUIDANCE. *Charles E. Limp* 530

A NOMOGRAPH FOR ESTIMATING THE RELIABILITY OF A TEST IN ONE RANGE OF TALENT WHEN ITS RELIABILITY IS KNOWN IN ANOTHER RANGE. *Edward E. Cureton and Jack W. Dunlap* 537

THE PROGNOSTIC VALUE OF THE LOW RANGE ARMY ALPHA SCORES. *Alfred H. Jones* 539

NOTES ON ARTICLES IN EDUCATIONAL PSYCHOLOGY IN CURRENT ISSUES OF OTHER MAGAZINES 542

NEW PUBLICATIONS IN EDUCATIONAL PSYCHOLOGY AND RELATED FIELDS OF EDUCATION 552

CORRECTIONS 560

Published Monthly Except June to August by

Warwick & York, Inc.,

10 East Centre Street — **Baltimore, Md.**

Entered as Second Class matter November 15, 1921, at the Post Office at Baltimore, Md., under the Act of March 3, 1879; additional entry as second class matter at York, Pa.

The Journal of Educational Psychology

is published monthly, by Warwick & York, Inc., Baltimore, Md., and York, Pa. Title page and index are bound in December number.

Manuscripts for publication, books or other materials for review, and news items should be addressed to Harold Rugg, Chairman of the Editorial Board, 400 W. 118th Street, New York City.

The price of the Journal is $4.00 a year, payable in advance. Foreign postage is 30 cents extra. Single issues are 60 cents each, and less than a full year's subscription is charged at single issue rate for every month ordered.

Subscribers should notify the publishers of change in address at least four weeks before publication of the issue with which change is to take effect. No claim for nonreceipt of an issue will be entertained unless made within two weeks after receipt of the next succeeding issue.

Warwick & York, Inc.,
Baltimore, Md.

TEACHERS' HEADQUARTERS

European
Plan

Entirely
Fire-Proof

Centrally
Located

EDWARD DAVIS
Manager

HOTEL RENNERT
BALTIMORE, MARYLAND

VOCATIONAL EDUCATION OF JUNIOR COLLEGE GRADE

By G. Vernon Bennett

This book is based upon a study recently made of the problem of vocational education of junior college grade. The study was prompted by leaders of the vocational education movement who believe that there exists a field that has been overlooked in the past, a plane of occupations lying between the professions and the artisan's trade. Three phases of the problem are examined: (a) The need for vocational education of junior college grade, especially under public auspices. (b) The present status of such education as it is attempted by factory schools, by systems of so-called "training on the job," and by schools conducted for profit. (c) A survey of present publicly supported educational agencies which might possibly be drafted for the task of carrying on vocational education of junior college grade if a nation-wide system should be organized.

Price $2.75 + 10¢ post.

WARWICK & YORK, Inc., *Baltimore, Md.*

THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume XX | October, 1929 | Number 7

CHOICE OF ITEMS FOR A TEST ON THE BASIS OF DIFFICULTY

PERCIVAL M. SYMONDS

Teachers College, Columbia University

One of the perplexing unsolved problems in connection with test construction concerns the choice of items for a test with reference to their difficulty. Should tests be scaled in difficulty, and if so from how easy to how hard should the items range? Should the test be steeply scaled or should the items have approximately the same level of difficulty? These questions are real and pertinent as tests become refined in their construction. It is the purpose of this paper to explore this problem and to lay down certain rules that can guide test makers in the selection of items on the basis of difficulty.[1]

Long ago (1916) Otis in studying the reliability of the Starch, Ayres, and Buckingham Spelling Scales gave us the law that the most reliable test is the one in which the average score made on the test by a given group is fifty per cent of the total number of items in the test.

In his original work on his Spelling Scale, Ayres indicated that the words of a given degree of spelling difficulty are spelled by fewer and fewer per cent of pupils in successive higher grades and that a given grade spells fewer and fewer per cent of words correctly at successive levels of spelling difficulty. Thorndike in "The Measurement of Intelligence" has investigated this whole problem and tells us that this same law is also true for individuals. That is, an individual is able to do a fewer per cent of tasks as the tasks ascend in difficulty. This relationship is expressed mathematically by the ogive curve, Figure 1.

[1] It is assumed that all the test items under consideration are equally valid and differ only in difficulty.

Thorndike has arbitrarily defined an individual's level or altitude of ability as that altitude at which he can perform successfully fifty per cent of the tasks.

Thorndike has also demonstrated two propositions that enable us to proceed with the problem without making unfounded assumptions. One is that the ogive curves for homogeneous material at successive levels of difficulty are identical in shape; *i.e.*, may be derived

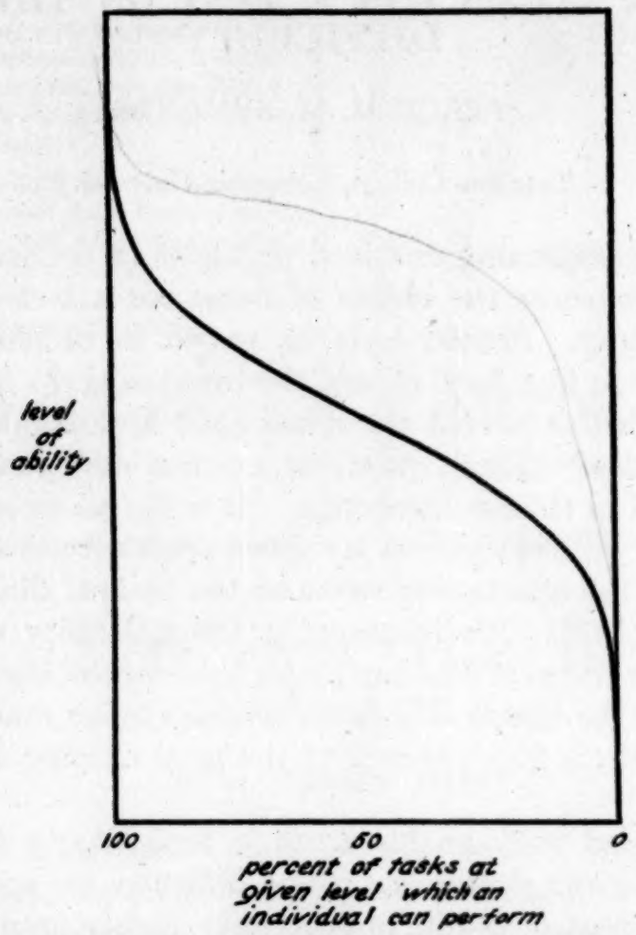


FIG. 1.—Ogive showing relation between ability and correctness of performance for test items of a given difficulty.

from distributions in terms of level having the same standard deviation. The other proposition is that the items at any level are a valid measure of the individual's ability at all levels in tasks homogeneous in construction and type of material.

With these two propositions we may envisage ability in terms of the graph, Figure 2. In this graph the abscissa represents the percent of items correct. The ordinate represents the level or altitude of difficulty on some absolute scale representing equal increments of

difficulty. The ogives represent test items of a given level of difficulty, and indicate the relation between level and per cent correct for items of a given difficulty.

To be more concrete let us assume that the horizontal line at level 0 represents the level of a pupil with seventh grade ability in spelling. Then ogive A will represent column X on the Ayres Spelling

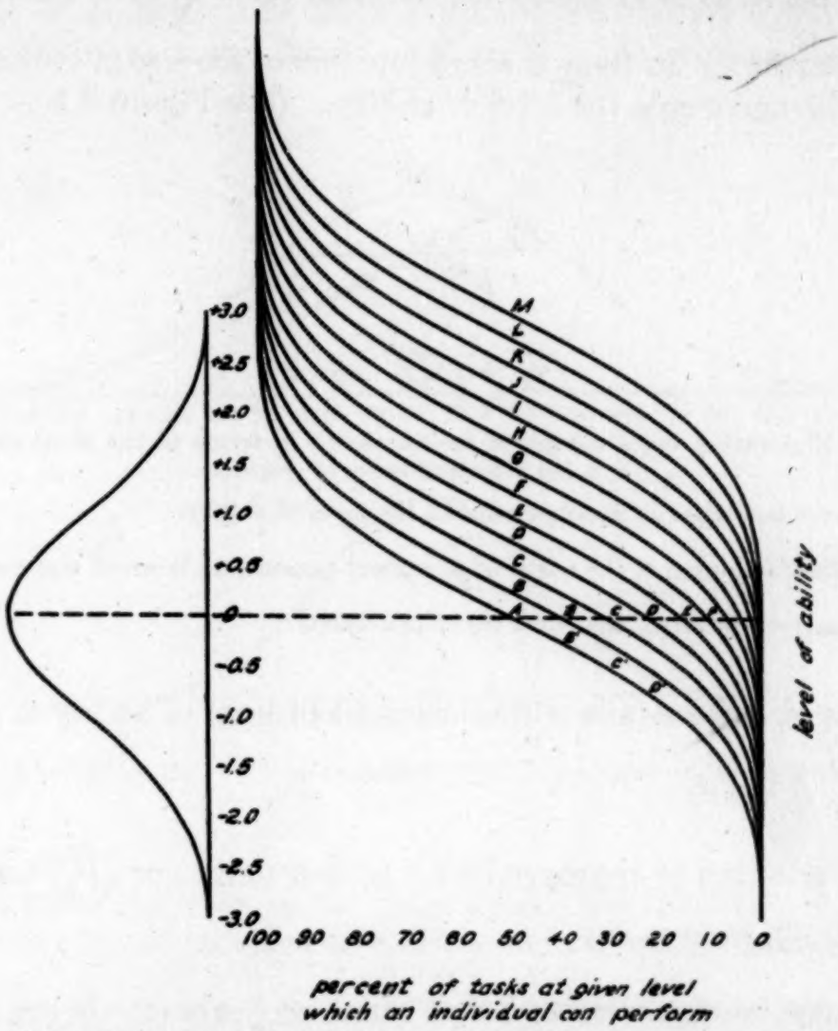
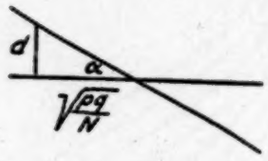


FIG. 2.—Family of ogives representing items of different difficulty showing relationship between ability and correctness of performance. The frequency distribution corresponding to ogive *A* is shown at the left.

Scale, the column containing words that can be answered with fifty per cent correctness by the average seventh grade child. And ogive B will stand for a list of words,—(say list Y on the Ayres Spelling Scale) the next degree harder in difficulty than ogive A. These words, being more difficult, will be spelled with a lower per cent of correctness by this individual—say forty per cent—as may be read from the graph.

We may define the *appropriateness* of an item by the accuracy with which it measures the ability of an individual. In the first place the accuracy of placement by an item is a function of reliability of the response to that item. If an individual answers N items all of a given degree of difficulty with p per cent accuracy the reliability of this per cent is given by the formula $\sigma p = \sqrt{\frac{pq}{N}}$. The accuracy of placement by an item is also a function of the size of the angle with which the ogive cuts the level of ability. (See Figure 3.)



$$\sigma_d = \sqrt{\frac{pq}{N}} \tan \alpha$$

FIG. 3.—Illustrating the accuracy of measurement in terms of the slope of the ogive and the standard error of response.

σ_d = standard error of measurement of the level of ability.

$\sqrt{\frac{pq}{N}}$ = standard error of the per cent of correct response to items of a given difficulty.

α = angle with which the ogive cuts the abscissa.

That is, the amount of displacement in level of ability is given by σ_d or $\sqrt{\frac{pq}{N}} \tan \alpha$.

Our criterion of appropriateness is then that σ_d or $\sqrt{\frac{pq}{N}} \tan \alpha$ be as small as possible.

It may be demonstrated that $\tan \alpha = \frac{1}{z}$ where z is the ordinate of the corresponding normal probability curve. (The ogive is the curve of the integral of the normal probability curve. The derivative of the ogive is correspondingly equivalent to the normal probability curve itself.) Our criterion of appropriateness may also be that $\frac{1}{\sigma_d}$ or $\frac{z}{\sqrt{pq}}$ be as large as possible. The problem then becomes that of determining for what values $\frac{z}{\sqrt{pq}}$ is a maximum. The function

$\frac{z}{\sqrt{pq}}$ when plotted bears resemblance in shape to the normal curve but approaches zero less slowly at the extremes. (See Figure 4.)

When we speak of the normal curve corresponding to an ogive we naturally are thinking of some one ogive (as ogive A). But our problem concerns the slopes with which the whole family of ogives cuts one level of ability. The same normal curve serves our purpose, however, for ogive A cuts level $(-.25)$ at B′ with the same slope that ogive B cuts level 0 at B.

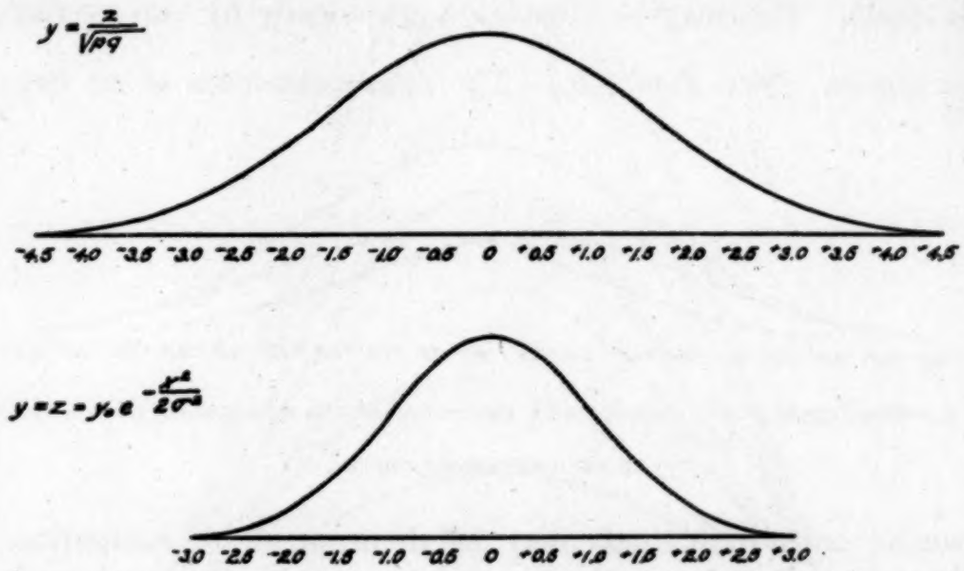


FIG. 4.—The curve $y = \frac{z}{\sqrt{pq}}$ compared with the normal probability curve.

First take the case of an individual. The problem is that of determining for what value of x of the normal curve in Figure 2 the expression $\frac{z}{\sqrt{pq}}$ is a maximum. This function is a maximum when $x = 0$ as determined by plotting the function from tables. (See Figure 4.)

Proposition 1.—The items with which one can measure the ability of an individual most accurately are the items that he can do with fifty per cent accuracy.

It is also evident that if one takes the sum of ten, forty or one hundred ordinates of the curve $\frac{z}{\sqrt{pq}}$ the maximum sum is obtained

by adding ordinates which lie at the point of origin of the normal probability curve.

Proposition 2.—The test which measures an individual most accurately is one made up of items all of which have a difficulty which the individual can solve with fifty per cent accuracy.

When two or more individuals are to be measured with the same set of items the problem becomes considerably complicated. To envisage the nature of the problem let us consider the case of two individuals, at levels 0σ and 1.0σ (A and E) on the chart. In this case we are concerned with the slope with which every ogive cuts both levels. This may be illustrated graphically by two overlapping $\frac{z}{\sqrt{pq}}$ curves. (See Figure 5.) The appropriateness of an item for

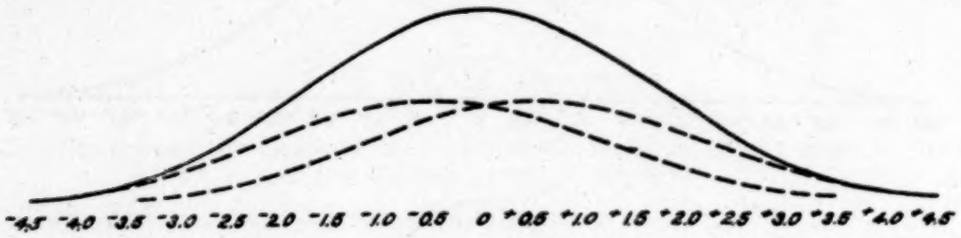


FIG. 5.—Overlapping $\frac{z}{\sqrt{pq}}$ curves and the curve of the summation of the ordinates of the component curves.

measuring both individuals may be determined by comparing the heights of overlapping ordinates and by noting the curve made by summing the ordinates of the two curves.[1] This latter curve has its maximum at the point midway between the points of origin of the separate curves.

Proposition 3.—The best item for measuring two individuals is the item lying in difficulty midway between the difficulty of the two items which can be answered with fifty per cent correctness by each of the individuals.

The sum of several ordinates is also a maximum when this same maximum ordinate is chosen.

[1] It becomes a question as to whether the sums of the ordinates should be a maximum, or the sums of the squares of the ordinates. It has been demonstrated that the sum of squares is a more reliable unit than the sum of the measures themselves but the difference in reliability is not great. Merely summing the ordinates in the present problem was used because it was more convenient to handle graphically and it is believed that the conclusions would not be changed by taking the sum of squares of the ordinates instead.

Proposition 4.—The best test for measuring two individuals is composed of items all with a difficulty which lies midway between the difficulty of the two items which can be answered with fifty per cent correctness by each of the individuals.

This is true for the case given where the two individuals are separated by 1σ in ability. If the two $\frac{z}{\sqrt{pq}}$ curves are widely

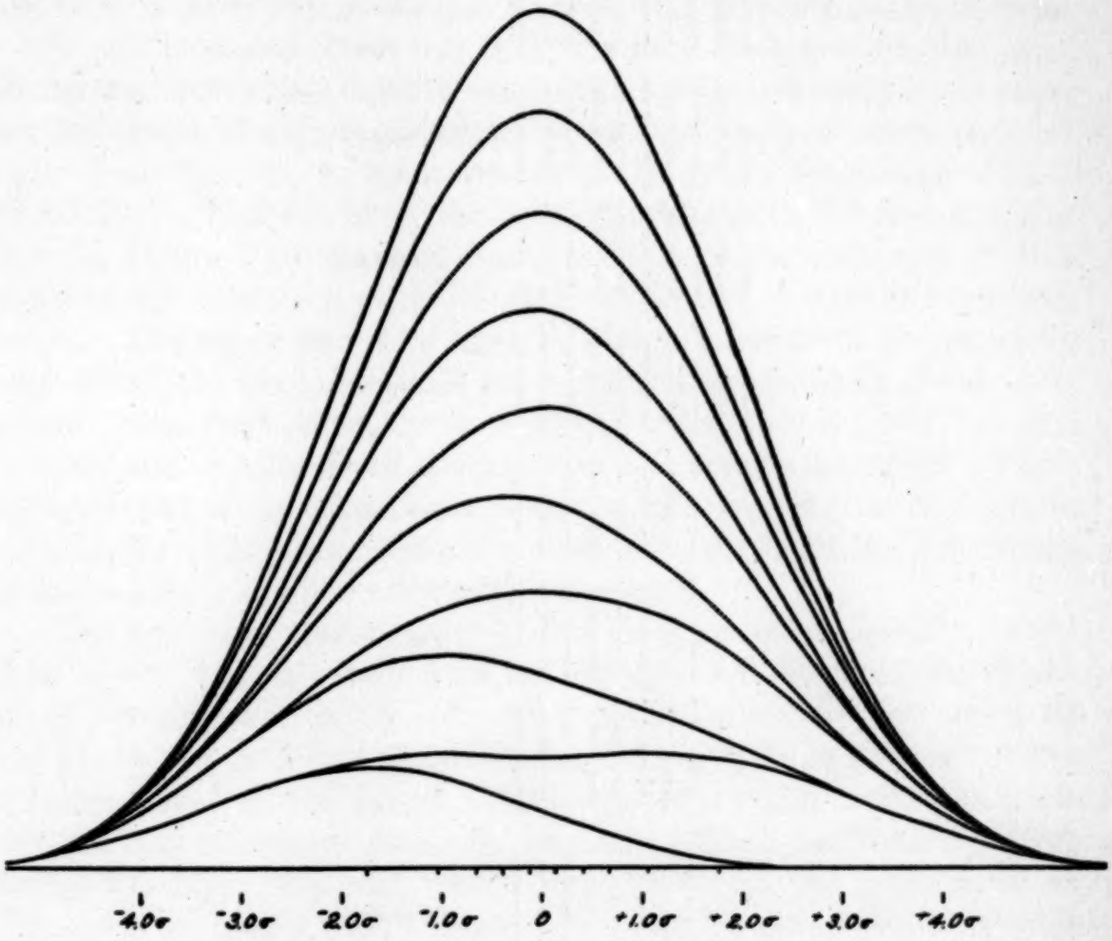


FIG. 6.—Summation of ten $\frac{\text{Ⅱ}z}{\sqrt{pq}}$ curves.

separated the composite curve will show two modes corresponding to the means of the two $\frac{z}{\sqrt{pq}}$ curves. However this case is so extreme that it would not ordinarily occur in the testing of a homogeneous school group. It would mean that if a child and an adult, widely separated in ability, were to be tested with the same test it would be better to select half of the items as most suitable for one, and

half most suitable for the other rather than to test both individuals with items midway between their abilities.

The case of choosing items to test ten individuals has been studied graphically. The ten individuals were chosen at intervals of $+1.755\sigma$,

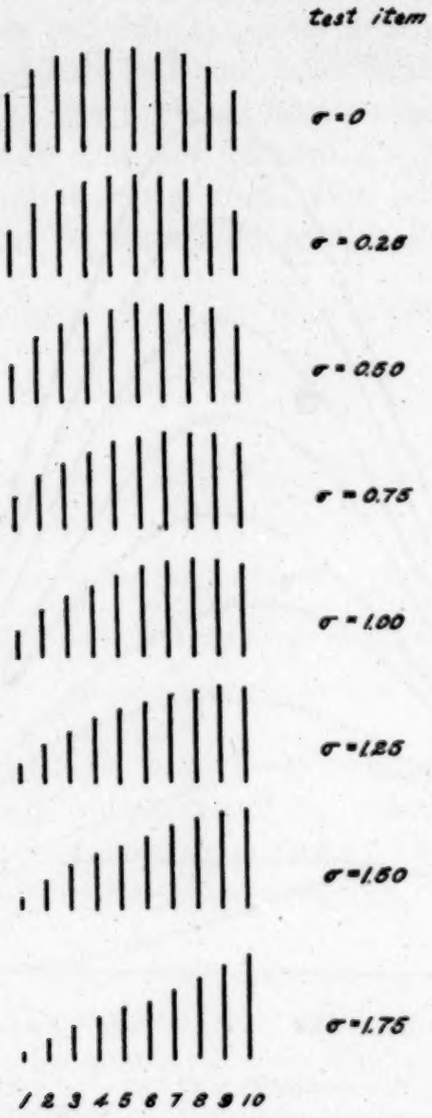


FIG. 7.—Accuracy with which ten individuals are measured by items of various difficulties. The ten lines in any group represent the accuracies with which the ten individuals are measured. Each row represents a test item of a given difficulty.

$+1.045\sigma$, $+.677\sigma$, $+.386\sigma$, $+.126\sigma$, $-.126\sigma$, $-.386\sigma$, $-.677\sigma$, -1.045σ and -1.755σ from a mean value to represent a group spread out according to the normal distribution of frequency. Ten curves of the type $\frac{z}{\sqrt{pq}}$ were piled one on top of the other in cumulative fashion

with the mean of each curve occurring at one of the points mentioned. (See Figure 6.) The resulting cumulative curve bears superficial resemblance to the normal probability curve. Its mode is at the mean of the means of all of the separate $\frac{z}{\sqrt{pq}}$ curves of which it is made up. This indicates that the most reliable single item with which to measure a group of individuals distributed normally is an item which the average individual of the group can answer with fifty per cent accuracy.

It will be noted, however, that this item does not measure each of the ten individuals equally accurately. The two individuals nearest the mean of all are measured most accurately whereas the two individuals (curves) at the extremes of the group are measured least accurately. This has been displayed graphically in the first group of lines in Figure 7 by making the lines equal to the ordinates of each of the component curves that lie on the ordinate $\sigma = 0$ of the composite curve. The other groups of lines in Figure 7 represent the accuracy with which the ten individuals are measured by ordinates of the composite curve (test items) lying $+.25\sigma$, $+.50\sigma$, $.75\sigma$, $+1.00\sigma$, $+1.25\sigma$, $+1.50\sigma$ and $+1.75\sigma$ from the mean of the composite curve. These indicate that as one takes items more and more difficult the individuals with higher ability are more adequately measured and the individuals of lower ability are less adequately measured.

The one item that measures the group most accurately is the item $x = 0$, but this item does not measure all the members of the group equally accurately. In fact individuals near the mean of the group are measured almost twice as reliably as individuals at the extreme edges of the group. Although one would build the most reliable test of several items by choosing all the items of the same mean difficulty, this test would measure all of the individuals of the group unequally well. Some individuals would be measured with greater reliability than other individuals. One has to choose then between constructing a test having the greatest possible reliability and a test in which all individuals are measured equally well. It is assumed that most teachers would prefer a test which would measure all of the pupils in a class equally reliably provided there is not at the same time too great a loss in total reliability.

To explore the possibilities of constructing a test in which all the ten individuals are measured equally reliably, combinations of two items were assembled and the lines in Figure 8 indicate how well they measure the ten individuals in the group. These groups of items

are 0 and 0; $+.25\sigma$ and $-.25\sigma$; $+.50\sigma$ and $-.50\sigma$; $+.75\sigma$ and $-.75\sigma$; $+1.00\sigma$ and -1.00σ; $+1.25\sigma$ and -1.25σ; $+1.50\sigma$ and -1.50σ and $+1.75\sigma$ and -1.75σ. In the last case all of the individuals are measured approximately equally reliably. In this particular case, then, the test that gives the most reliable measure of all individuals and

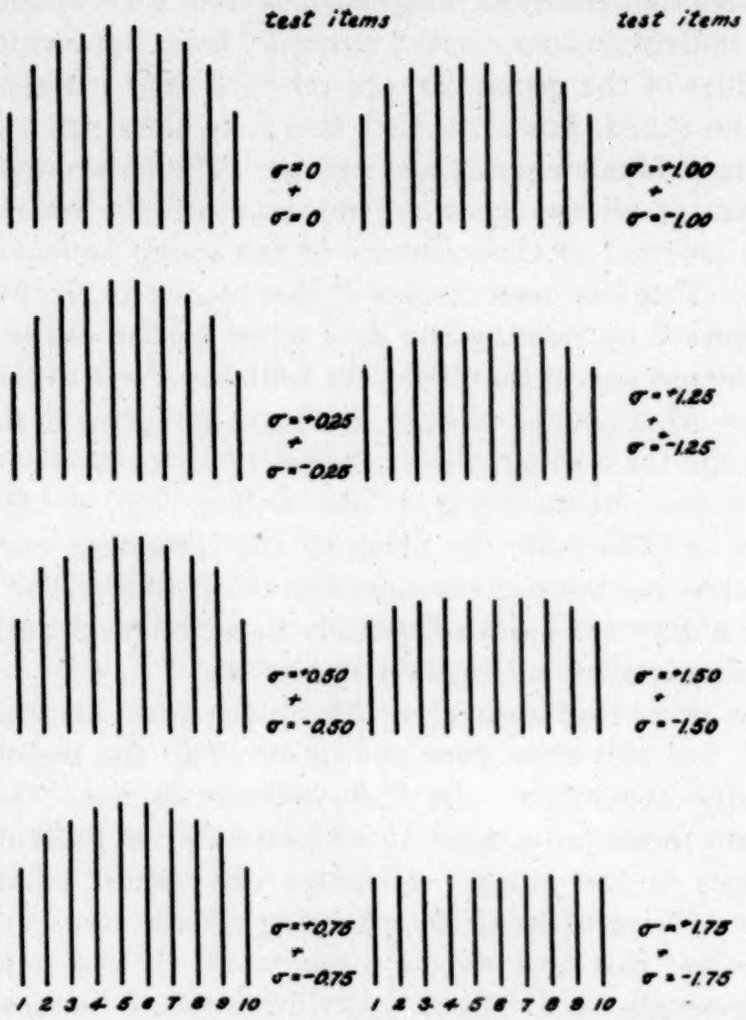


FIG. 8.—Accuracy with which ten individuals are measured by two items taken at equal distances on either side of the median difficulty. The ten lines in any group represent the ten individuals measured. Each group represents the combination of two test items.

that at the same time measures all individuals equally reliably is a test in which half of the items are chosen $+1.75\sigma$ above the average difficulty of the group and the other half -1.75σ below the average of the group. One should note, however, that the extreme individuals do not gain appreciable reliability by selecting the pair of $+1.75\sigma$, -1.75σ items over what they have with a pair of 0σ, 0σ items while

the rest of the items lose considerable reliability. All in all, then, the best test for this particular group of ten individuals is that which contains items all of the same optimum median difficulty.

Were the individuals to string out beyond 1.75σ the situation would change and the extreme individuals would indeed suffer from poor measurement by the median item. To determine how far individuals might be expected to range in any group, the Thorndike CAVD scale and Ayres Spelling Scale were studied from this point of view.

The Thorndike CAVD scale was so constructed that one unit on the scale is the equivalent of one standard deviation in a ninth grade group. Six SD, CAVD, is taken as the range for a group. The ogives of items of equal difficulty have a corresponding SD value of approximately 2.65 on the CAVD scale. Dividing 6 by 2.65 we obtain 2.26SD as the probable extreme difference between individuals in an ordinary school class on the $\frac{z}{\sqrt{pq}}$ distribution, or individuals at $+1.13\sigma$ and -1.13σ respectively. Another computation on the CAVD scale using a different set of initial values gives $+1.55\sigma$ and -1.55σ as the probable σ values of the extreme individuals. Similar calculations on the Ayres Spelling Scale place the extreme individuals at $+1.62\sigma$ and -1.62σ. All of these values are well within the $+1.75\sigma$ and -1.75σ of our extreme individuals.

Proposition 5.—The best test for measuring a typical school grade or class is a test in which all of the items have a difficulty such that they can be answered with fifty per cent accuracy by the average individual of the group.

When a standardized test is constructed for use over several grades the above rule does not hold. In this case the extreme individuals in the upper and lower groups would be inadequately tested and items should be chosen over a range of difficulty.

Proposition 6.—The best test designed to measure several consecutive grades or classes is one in which the items have been so selected that they range evenly in difficulty from the level of difficulty which can be done with fifty per cent accuracy by the average member of the lowest group to be tested to the level of difficulty which can be done with fifty per cent accuracy by the average member of the highest group to be tested.

Looking into the future the line of development which promises most for the testing movement would seem to be the construction of

lists of valid items with difficulty values attached for each function and each subject to be tested. Scales or lists of items of this kind are already available in such subjects as spelling (the Ayres Spelling Scale), geography (the Hahn-Lackey Geography Scale), and shorthand (The Hoke Scale of Shorthand Items). Thorndike has prepared but has not published lists with difficulty values for vocabulary, language completion and arithmetic problems.

With such lists teachers could make up their own tests and would not be dependent on using tests printed and sold by commercial publishing houses. There are two advantages in this. One is the saving in cost of the material. A teacher, having selected the items with which she wishes to test, could have them mimeographed or printed in any convenient way. The common use of duplicating or printing machines today makes this method perfectly feasible. Or a superintendent could have his tests locally assembled to meet his own needs in the best way, and could have them printed locally and probably more cheaply than they can now be purchased from the book companies.

Tests thus locally constructed for specific grades or classes would be so far superior to the standardized tests now available that the same reliability which is now obtainable with standardized tests could be obtained in less testing time. That is, standardized tests which have been constructed to cover the needs of several grades contain "dead" material for the average class—material that is too hard or too easy. On the high school level a standardized test which covers the whole year's work is wasteful. If given near the beginning of the year much of the difficult material which belongs in the latter part of the course will not be touched and if given near the end of the year much of the material is so easy as to fail to "test."

Instead of trying to measure a group with limited distribution of ability with a scale running the total range of ability, one selects material for a test all of approximately the same difficulty which corresponds to the average level of ability of the group to be measured.

In closing, the directions for assembling a test will be repeated, assuming that one has a list of valid test items with difficulty values in terms of grades, ages, or months a subject has been studied. First, estimate informally the general level of ability of the class. In the elementary school, the *grade* may be used with perhaps a minor correction if the class is inferior or superior to the grade norms. In high school if it is a foreign language, one knows how many years

the language has been studied and this may be used as a preliminary estimate. If it is a science, the estimate may be in terms of number of months the subject has been studied. After the average achievement of the group has been estimated from such ready criteria as are available, select as many test items as are desired to give the requisite reliability or to occupy the desired testing time. Select these items as nearly as possible of the same difficulty, which can be answered by the average pupil with fifty per cent accuracy as estimated on the basis of the preliminary estimate of the advancement of the group.

THE SUBJECT-MATTER PROGRESS OF THREE ACTIVITY SCHOOLS IN HAWAII, WITH A NOTE ON STATISTICAL TECHNIQUE

HELEN G. PRATT

Division of Research, Department of Public Instruction of Hawaii

JACK W. DUNLAP

Territorial Normal and Training School, Honolulu, Hawaii

AND

EDWARD E. CURETON

Territorial Normal and Training School, Honolulu, Hawaii

I. The Problem

Whenever a school makes an attempt to break away from tradition and adopt a new procedure, conventional school people question its efficiency. A progressive public school cannot ignore this questioning. The present paper attempts to show in the case of three activity schools in Hawaii, whether subject-matter achievement as measured by the Stanford Achievement Test tends to improve, remain constant or decrease under such a program as compared to the traditional program of studies.

During the school year 1927–1928, three schools in Hawaii were engaged in working out an activity program. Other schools changed their procedure too late in the school year to be included in this study. Kawananakoa, the experimental school directly under the Division of Research, and located in Honolulu, had a somewhat selected faculty, though all the members were in varying degrees inexperienced in the technique of the activity school and in some cases were unacquainted with the theory. The principal of the school conducted throughout the year an intensive course in activity school theory, and all his teachers were enrolled. The activities of this school were much restricted by the fact that it was obliged to hold its sessions in a Japanese language school. No materials or equipment of any kind could be left in the school rooms, and until late in the year there were not even storage closets in which things could be placed. For many weeks the activities in the school were limited to those which could be carried on through the use of materials which could be taken home each night by the teachers and pupils. Neither teachers nor pupils could remain after school for any purpose, as the sessions of the lan-

guage school began as soon as the public school closed. The school plant and surroundings were inferior to those of the average public school in Hawaii. The children of this school were tested with the Stanford Achievement Test, Form A, in April, 1927; some entering children were tested with the same form of the test in September, 1927; and all were tested again in May, 1928 with Form B. The tests were given to all the children from Grades II to VIII inclusive. The other two schools, Kamehameha III and Wailuku, are located on the island of Maui. Both were under new principals, who conducted throughout the year classes in the theory and practice of the activity school, which were attended by most of their teachers. The teachers in these schools were not selected, and the equipment and materials were those of the average public school in Hawaii. These schools were tested with the Stanford Achievement Test, Form A, in October, 1927. Each school had many over-age retarded children, who were reclassified by the Division of Research in November, 1927. They were moved up and grouped together.

The detailed procedure in the three schools varied, but in general each school showed these features: A deliberate attempt to find and work out activities which were real in the lives of the children, an increase in the types and variety of reading material used, informal procedure, within the limits of the bondage of the traditional desks, and an emphasis upon social rather than specific subject-matter values.

Teachers participated in the giving and scoring of the tests; but after instruction and under supervision. All scoring was checked, and all computations were checked twice. Securing correct chronological ages offers a peculiar problem in Hawaii. The Oriental calls a child a year old at birth. It is common for both parents to work away from home, so they strongly desire to get the children in school as early as possible. Therefore they have often in times past stated that a child was six when according to the Oriental calculation he was five, and according to ours he was four. More common still is the attempt to enter the child at what we call five. Methods and materials adapted to the child of six are not suited to the child of four or five; hence the child who is enrolled too young is apt to have a school history of failure and retardation. As a child gets older and is not promoted regularly, the tendency is to report him as younger than he is. If no check is made on the reported age, it may change several times during a school career. Schools now ask for presentation

of the birth certificate, but birth certificates may be obtained at any time, sometimes years later than the date of birth, on the statement of the parents. Besides these factors, there is the inaccuracy arising when numbers of people, many of whom have not had clerical training, are involved in the keeping of records. Within the limits of all these conditions, the ages of the children in the three schools were checked as closely as possible. In these schools, the checking went on for the whole period between the giving of the two tests, and by the time the second test was given the birth date and the calculation of the ages had been checked.

The problem then became that of determining whether the classes in these schools had made as much progress, as measured by the Stanford Achievement Test, or more or less than they might reasonably have been expected to make under the traditional program of studies. The usual method of attacking this problem is to compare the gain of the experimental groups with the gain of control groups carefully selected to have the same average ability and spread as the experimental groups at the beginning of the experiment. The present solution differs in that it takes as the control group the population used in standardizing the test; the averages for this control group then become the published norms of the test; and appropriate statistical methods must be devised to correct for the variations in means and variabilities between the groups. The essential assumption at the basis of the method here presented is that the real variability of a group does not change significantly over a period of seven to thirteen months; *i.e.*, that differences between the standard deviations of the first and second tests are due to changes in scale-steps, inequality of forms of the test and chance errors of measurement rather than to changes in the real variability of the group. On this assumption, the difference between the average score on the test and the age norm for the average chronological age, expressed in terms of the standard deviation of the group as the unit, should provide a measure of the achievement of the class independent of the units of measurement of the test: *i.e.*, a group standard score. If the efficiency of the instruction remains constant, this group standard score should also remain constant, and the difference between two such scores for tests given at the beginning and end respectively of a period of instruction should provide a measure of the efficiency of the instruction in developing the abilities measured by the tests. If the difference is zero the efficiency of the instruction may be presumed to be equal to the average

efficiency in the groups used in standardizing the test; or the group may be said to have made the amount of progress reasonably to be expected of such a group.

Essentially then, the problem became that of determining this difference and whether it differed significantly from zero. To do this it was necessary to determine its standard error.

II. Statistical Technique

Let

$$f_1 = \frac{M_1 - Z_1}{\sigma_1}, \quad f_2 = \frac{M_2 - Z_2}{\sigma_2}, \quad d = f_2 - f_1$$

where M_1 is the mean score at the first testing
M_2 is the mean score at the second testing
σ_1 is the standard deviation at the first testing
σ_2 is the standard deviation at the second testing
Z_1 is the age norm at the first testing
Z_2 is the age norm at the second testing
N is the population of the group
d is the measure of efficiency

We require to know the standard error of d. The standard error of any difference is

$$\sigma_d = \sqrt{\sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2 r_{12}}$$

which in this case becomes

$$\sigma_d = \sqrt{\sigma_{f_1}^2 + \sigma_{f_2}^2 - 2\sigma_{f_1}\sigma_{f_2} r_{f_1 f_2}} \tag{1}$$

We then require σ_{f_1}, σ_{f_2}, and $r_{f_1 f_2}$.

$$f_1 = \frac{M_1 - Z_1}{\sigma_1}$$

Assuming that the error in Z_1 is insignificant in comparison with those in M_1 and σ_1, we may treat Z_1 as a constant, whence, taking logarithmic differentials,

$$\frac{df_1}{f_1} = \frac{dM_1}{M_1 - Z_1} - \frac{d\sigma_1}{\sigma_1}$$

Squaring, summing and dividing by the theoretical infinite population,

$$\frac{\sigma_{f_1}^2}{f_1^2} = \frac{\sigma_{M_1}^2}{(M_1 - Z_1)^2} + \frac{\sigma_{\sigma_1}^2}{\sigma_1^2} - \frac{2\sigma_{M_1}\sigma_{\sigma_1} r_{M_1\sigma_1}}{\sigma_1(M_1 - Z_1)}$$

Since $r_{M_1\sigma_1} = 0$, the third term drops out, and we have

$$\sigma_{f_1}^2 = \left[\frac{\sigma_1^2\sigma_{M_1}^2 + \sigma_{\sigma_1}^2(M_1 - Z_1)^2}{\sigma_1^2(M_1 - Z_1)^2}\right]\frac{(M_1 - Z_1)^2}{\sigma_1^2}$$

Substituting the values of σ_{M_1} and σ_{σ_1} and reducing,

$$\sigma_{f_1}^2 = \frac{2\sigma_1^2 + (M_1 - Z_1)^2}{2N\sigma_1^2}$$

and

$$\sigma_{f_1} = \left[\frac{2\sigma_1^2 + (M_1 - Z_1)^2}{2N\sigma_1^2}\right]^{1/2} \tag{2}$$

Similarly

$$\sigma_{f_2} = \left[\frac{2\sigma_2^2 + (M_2 - Z_2)^2}{2N\sigma_2^2}\right]^{1/2} \tag{3}$$

Let M equal the theoretical infinite population. Then

$$r_{f_1f_2} = \frac{\frac{1}{M}\sum f_1f_2}{\sigma_{f_1}\sigma_{f_2}} = \frac{\frac{1}{M}\sum\left(\frac{M_1 - Z_1}{\sigma_1}\right)\left(\frac{M_2 - Z_2}{\sigma_2}\right)}{\sigma_{f_1}\sigma_{f_2}}$$

$$= \frac{\frac{1}{M}\sum(M_1M_2 - M_1Z_2 - Z_1M_2 + Z_1Z_2)}{\sigma_1\sigma_2\sigma_{f_1}\sigma_{f_2}}$$

$$= \frac{r_{M_1M_2}\sigma_{M_1}\sigma_{M_2} - r_{M_1Z_2}\sigma_{M_1}\sigma_{Z_2} - r_{Z_1M_2}\sigma_{Z_1}\sigma_{M_2} + z_1z_2\sigma_{Z_1}\sigma_{Z_2}}{\sigma_1\sigma_2\sigma_{f_1}\sigma_{f_2}}$$

Treating Z_1 and Z_2 as constants, and noting that the correlation between a variable and a constant or between two constants equals zero, and that the correlation between the means of two random samples equals the correlation between the corresponding variables, we have

$$r_{f_1f_2} = \frac{r_{12}\sigma_{M_1}\sigma_{M_2}}{\sigma_1\sigma_2\sigma_{f_1}\sigma_{f_2}}$$

Substituting the values of σ_{M_1} and σ_{M_2}

$$r_{f_1f_2} = \frac{r_{12}}{N\sigma_{f_1}\sigma_{f_2}} \tag{4}$$

Substituting in (1),

$$\sigma_d = \sqrt{\sigma_{f_1}^2 + \sigma_{f_2}^2 - \frac{2r_{12}\sigma_{f_1}\sigma_{f_2}}{N\sigma_{f_1}\sigma_{f_2}}} = \sqrt{\sigma_{f_1}^2 + \sigma_{f_2}^2 - \frac{2r_{12}}{N}} \qquad (5)$$

and

$$\mathrm{PE}_d = .6745\sqrt{\sigma_{f_1}^2 + \sigma_{f_2}^2 - \frac{2r_{12}}{N}} \qquad (6)$$

III. Results

The test records for the three schools were sorted by ages. Each age group in each school was studied separately. The mean composite score on the Stanford Achievement Test and the standard deviation were computed for each testing of each group. In determining the age norm, the median chronological age was used rather than the mean in order to avoid difficulties which might arise in taking the mean of ages not all of which were known to be entirely accurate. Age norms were taken from the Manual of Directions of the test. The mean score of each test for each group was compared with the corresponding norm, and the difference divided by the standard deviation of the group. The difference between the two group standard scores for each group was found, and the probable error of this difference was computed by formulas (2), (3) and (6).

After computing three of the intercorrelations (between Form A given at the beginning and Form B given at the end of the experiment), which came out .95, .96 and .95, the value .95 was substituted for r_{12} in all cases. This value is probably not significantly different from the true correlation in most cases; and where it is it has the effect of increasing the probable error of the difference slightly; so no difference will erroneously be estimated to be significant in any case.

The results are as follows:

N	Median CA	Gain or Loss
	Kawananakoa	
60	9- 7	Gain .096 ± .029
56	10- 7	Loss .111 ± .053
50	11- 6	Loss .040 ± .079
49	12- 6	Gain .012 ± .087
46	13- 5	Gain .20 ± .13
36	14- 4	Gain .15 ± .28
	Wailuku	
44	9- 1	Gain .113 ± .032
62	9-11	Gain .030 ± .037
85	10-10	Loss .100 ± .062
85	11-11	Gain .076 ± .073
55	12- 9	Gain .135 ± .081
41	13- 9	Gain .188 ± .099
30	14- 9	Gain .07 ± .19
	Kamehameha II	
41	9-1	Loss .020 ± .035
81	9-10	Loss .008 ± .027
102	10-10	Loss .001 ± .053
87	11-10	Gain .043 ± .055
60	12- 9	Gain .06 ± .10
46	13- 8	Gain .09 ± .12
24	14- 7	Gain .13 ± .22

These figures show that on the things which the Stanford Achievement Test measures, there was no significant loss, and in the case of the youngest groups in two of the schools, there was significant gain. All three schools maintained about the same rate of subject-matter progress under the new program as under the old.

A COMPARISON OF CERTAIN PHASES OF MUSICAL ABILITY OF COLORED AND WHITE PUBLIC SCHOOL PUPILS

C. T. GRAY

University of Texas

AND

C. W. BINGHAM

Junior College, South Park, Texas

The purpose of this study is to compare colored and white children with respect to certain aspects of musical ability. Based upon results from the Seashore Test of Musical Ability three different comparisons are set forth as follows: (1) Total scores for colored and white children are compared; (2) results for colored and white pupils who have had training in music are contrasted with those without training; and (3) scores for colored and white pupils are arranged so as to throw light on musical ability in its relation to chronological age, mental age, so-called intelligence and educational progress.

The pupils selected for the study were from the public schools of South Park, Beaumont, and Port Arthur, Texas. Both white and colored pupils were examined from the South Park and Beaumont Schools, while only colored pupils were examined from the Port Arthur Schools. Most of the pupils were from Grades VII and VIII, only a few from Grade VI being taken. The total number of pupils examined was two hundred fifty-eight colored and two hundred nineteen whites.

A distinction is made between full-blooded colored pupils and those having a mixture of white with colored blood. The former were classed as *negro*, whereas the latter were termed *mulatto*. The term *colored* is used to include both groups. Classification on this basis was made by the school principals.

By means of a questionnaire, the pupils were classified as trained or untrained. This classification was made on the basis of special musical instruction other than that received in school. If a pupil had had special instruction for one year or more with at least two lessons per week, he was placed in the trained group. Any pupil with less than one year of special instruction was classed as untrained. In this part of the investigation, six groups of pupils were considered—

The results are as follows:

N	Median CA	Gain or Loss
	Kawananakoa	
60	9- 7	Gain .096 ± .029
56	10- 7	Loss .111 ± .053
50	11- 6	Loss .040 ± .079
49	12- 6	Gain .012 ± .087
46	13- 5	Gain .20 ± .13
36	14- 4	Gain .15 ± .28
	Wailuku	
44	9- 1	Gain .113 ± .032
62	9-11	Gain .030 ± .037
85	10-10	Loss .100 ± .062
85	11-11	Gain .076 ± .073
55	12- 9	Gain .135 ± .081
41	13- 9	Gain .188 ± .099
30	14- 9	Gain .07 ± .19
	Kamehameha II	
41	9-1	Loss .020 ± .035
81	9-10	Loss .008 ± .027
102	10-10	Loss .001 ± .053
87	11-10	Gain .043 ± .055
60	12- 9	Gain .06 ± .10
46	13- 8	Gain .09 ± .12
24	14- 7	Gain .13 ± .22

These figures show that on the things which the Stanford Achievement Test measures, there was no significant loss, and in the case of the youngest groups in two of the schools, there was significant gain. All three schools maintained about the same rate of subject-matter progress under the new program as under the old.

A COMPARISON OF CERTAIN PHASES OF MUSICAL ABILITY OF COLORED AND WHITE PUBLIC SCHOOL PUPILS

C. T. GRAY

University of Texas

AND

C. W. BINGHAM

Junior College, South Park, Texas

The purpose of this study is to compare colored and white children with respect to certain aspects of musical ability. Based upon results from the Seashore Test of Musical Ability three different comparisons are set forth as follows: (1) Total scores for colored and white children are compared; (2) results for colored and white pupils who have had training in music are contrasted with those without training; and (3) scores for colored and white pupils are arranged so as to throw light on musical ability in its relation to chronological age, mental age, so-called intelligence and educational progress.

The pupils selected for the study were from the public schools of South Park, Beaumont, and Port Arthur, Texas. Both white and colored pupils were examined from the South Park and Beaumont Schools, while only colored pupils were examined from the Port Arthur Schools. Most of the pupils were from Grades VII and VIII, only a few from Grade VI being taken. The total number of pupils examined was two hundred fifty-eight colored and two hundred nineteen whites.

A distinction is made between full-blooded colored pupils and those having a mixture of white with colored blood. The former were classed as *negro*, whereas the latter were termed *mulatto*. The term *colored* is used to include both groups. Classification on this basis was made by the school principals.

By means of a questionnaire, the pupils were classified as trained or untrained. This classification was made on the basis of special musical instruction other than that received in school. If a pupil had had special instruction for one year or more with at least two lessons per week, he was placed in the trained group. Any pupil with less than one year of special instruction was classed as untrained. In this part of the investigation, six groups of pupils were considered—

negroes, trained and untrained; mulattoes, trained and untrained; whites trained and untrained.

The tests for musical ability were given according to the printed directions to pupils in groups of fifty during the morning session of school between the hours of ten and twelve. All tests were given and scored by the same person. The Otis Group Intelligence Scale, Advanced Examination, Form A, was used as a measure of intelligence, and educational standing was determined by means of the Stanford Achievement Test, Form A. These tests were given a few days after the Seashore tests were given.

The results of the music tests in terms of medians and sigmas of differences are shown in Table I.

TABLE I.—SHOWING THE MEDIANS AND SIGMAS OF DIFFERENCES FOR SCORES MADE BY COLORED AND WHITE PUPILS

	Pitch	Intensity	Time	Consonance	Memory
Medians for colored..	56.1	69.2	65.7	32.5	21.2
Medians for white...	64.7	80.1	70.2	31.9	29.4
Sigmas of differences.	1.21	1.34	.89	.47	.79

From this table it is clear that the median score for white pupils exceeds that for the colored children in all tests with the exception of consonance. In this, the colored pupils have a median of 32.5 while the white pupils have one of 31.9, a difference of only .6. All of the differences seem valid except that for consonance and here the chances are four out of five that the difference would be greater than zero in favor of the colored pupils. These results are directly opposed to the popular belief concerning the relative musical ability of the two races.

RESULTS OF COMPARING THE TRAINED AND UNTRAINED GROUPS

The two hundred fifty-eight negroes and two hundred nineteen whites were divided into six groups, as already described, according to training, which was determined by the questionnaire previously mentioned. Scores for these different groups are shown in Table II. Unfortunately the results for consonance were lost so that no data are shown for this test. Columns 1, 3, 5, and 7 indicate differences between the medians for the various groups, while columns 2, 4, 6 and 8 show the sigmas for the differences. The letters C, M, and W stand colored, mulatto, and white respectively. The table may be read as follows:

In the untrained group the difference in the medians for the test in pitch for the colored and mulatto group is 2.6, and the sigma of this difference is .80. The negative mark before the difference means that the colored pupils for whom results are represented in the second of the two columns which designate subjects are superior to the mulattoes. If no sign had appeared here, it would have meant that the mulattoes were superior to the negroes.

TABLE II.—SHOWING DIFFERENCES IN TRAINED AND UNTRAINED GROUPS

	Pitch		Intensity		Memory		Time	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Untrained								
M-C	−2.6	.80	6.9	.29	−3.2	.4	5.4	.72
W-C	−2.9	.79	20.0	1.01	1.4	1.05	.9	.41
W-M	−.3	.26	13.1	.98	4.6	.92	−4.5	.13
Trained								
M-C	3.8	.61	14.7	.59	9.8	.24	0	.51
W-C	18.3	.28	29.7	.40	14.9	.26	10.7	.33
W-M	14.5	.58	6.0	.44	5.0	.36	10.7	.45
Trained and Untrained								
C-C	4.1	.2	4.0	.40	.7	.59	−.3	.31
M-C	7.9	.57	18.7	.45	10.6	.64	−.3	.42
W-C	22.4	.18	24.7	.16	15.6	.55	10.4	.16
M-M	10.5	.59	11.8	.52	13.8	1.37	−5.7	.47
W-M	25.0	.26	17.8	.30	18.8	.19	5.0	.25
C-M	6.7	.30	− 2.9	.47	3.9	.28	−5.7	.37
C-W	18.3	.26	−20.7	1.03	−14.9	.94	−10.7	.32
M-W	14.5	.56	− 1.3	1.05	9.2	.97	4.5	.42
W-W	25.3	.20	− 4.7	.7	14.2	.92	9.5	.18

The results for the untrained groups set forth eight comparisons involving negroes, eight involving mulattoes, and eight involving whites. In these comparisons, the negroes are superior in three cases these being in the results for pitch and memory. In the same manner, the mulattoes are superior in four instances. The difference in the results for the test of pitch between mulattoes and whites is .3 and the sigma of the difference indicates that it is not reliable. Consideration of the results of the white shows them to be superior in

five cases. The implication of the results for the entire group of untrained subjects is that there is little difference in the ability of negroes and mulattoes, and that the whites are slightly superior to other groups.

In the trained group the results show the negroes to be superior in no comparison, the mulattoes in four comparisons, and the whites in eight comparisons. These data indicate that training is more effective in the case of mulattoes than in the case of negroes, and more effective for whites than for either negroes or mulattoes.

Twelve comparisons are shown for each group in the results for the trained and untrained groups. In these comparisons, the untrained negroes are superior in only two cases. These are in time, and here the differences are small and unreliable. In the same manner the untrained mulattoes are superior in three comparisons. Two of these are in time and one in intensity. In each instance the differences seem reliable. Finally, the untrained whites are superior in five cases. Three of these are in the comparisons with negroes, and one with each of the other groups. In all cases except one the differences seem reliable. From the results for the trained versus the untrained groups three conclusions may be drawn as follows: First, the effectiveness of training is shown in two phases of musical ability, pitch and memory. Second, the superiority of the untrained groups shows itself for the most part in the results for intensity and time. From this it seems that these phases of musical ability are less subject to training than the other factors. Third, the whites appear superior to the negroes and mulattoes, and the mulattoes to the negroes.

Results for Other Test.—The pupils selected for this part of the study included twenty-seven negro boys, thirty-one negro girls, forty-nine white boys, and sixty-three white girls. Both the negro and white pupils were selected from schools in which music has not been stressed. Very few of the pupils have had a good musical environment, and few have had special instruction in music. The groups were nearly equal in musical training. Regular academic work, in the two schools as measured by Standard Tests compares well with that of the average school.

The educational age, the mental age, and the index of brightness were determined by the tests previously mentioned. The scores on the different elements of the music tests were added together and recorded as total music score (T. M. S.) for each pupil. The results are given in Table III.

TABLE III

		COLORED	WHITE
CA	median	160	164
	Sigma	13.70	12.28
	Sigma of Difference, 18.39		
MA	median	120	178
	Sigma	16.27	29.28
	Sigma of Difference, 33.49		
EA	median	133	160
	Sigma	10.99	13.89
	Sigma of Difference, 17.43		
IB	median	76	108
	Sigma	11.42	22.67
	Sigma of Difference, 25.36		
TMS	median	222	260
	Sigma	24.8	32.21
	Sigma of Difference, 40.64		

From the above table, the median chronological age for the colored pupils is a few months less than for the white. The mental age of the colored pupils is seen to be almost three years below the mental age of the whites. The index of brightness for the colored pupils is seventy-six, while for the white pupils it is one hundred eight. The educational age of the colored pupils is almost the same as their mental age. The educational age for the white pupils is less than their mental age and almost identical with their chronological ages.

The medians for the last three measures indicate superiority on the part of the whites but the sigmas of the differences shows reliability only in the case of the index of brightness. The sigmas of the distribution point to greater variability of the same group. Various correlations between the different measures were determined. The results, however, except in one case are of little significance inasmuch as only a small positive correlation was found to exist.

A correlation of interest seems to exist between index of brightness and total music score. This is shown in Table IV.

TABLE IV.—SHOWING THE CORRELATION BETWEEN INDEX OF BRIGHTNESS AND TOTAL MUSIC SCORE

Colored boys	.580	+−.085
Colored girls	.534	± .111
White boys	.700	± .051
White girls	.677	± .048

The higher correlation in the case of the whites is probably due to the fact that musical ability as found in their case seems more amen-

able to training than in the case of the colored group. The difference in the correlations for boys and girls is difficult to account for.

In conclusion, four points may be made as follows:

1. The superiority of the whites when compared with either mulattoes or negroes seems apparent in most comparisons made.

2. The superiority of the mulatto to the negro is apparent when training has been applied.

3. Those phases of musical ability which are least amenable to training are intensity and time.

4. A relative high correlation is found to exist between total musical score and index of brightness.

THE RELATIONSHIP OF MOTOR ABILITY TO THE INTELLIGENCE OF KINDERGARTEN CHILDREN

OSCAR E. HERTZBERG

Colorado State Teachers College

A casual observation of the activities of young children reveals marked differences in motor aptitudes. For example, some children are skillful in feeding themselves at an early age while others seem not to have acquired dexterity even at the time of entering school. There have come under the observation of the writer a number of children in the kindergarten and Grade I who, although otherwise bright, were extremely awkward in their attempts to handle a pair of scissors or a saw. How much these differences may be due to heredity, how much to early environment, how much to training, to physical well-being, to temperament as revealed in desire and interest, or to general mental development we cannot say. Experimental studies indicate that desirable traits are related, and that individuals superior in one ability tend to be superior in other abilities, and *vice versa*. Certain it is that infants and very young children respond to their environment in great part through motor reactions, and, this being true, about the only check that can be obtained on the general development of the very young child is through a study of these motor reactions. And yet common observation and experimental data concerning the relationship between general intelligence and manual dexterity for older children and adults indicate that there comes a time in the development of the individual when abstract intelligence is not revealed even approximately through motor dexterity. Previous to this study, uncontrolled observations had led the writer to believe that this latter condition had arrived at the kindergarten period, otherwise how explain the seemingly manifest differences in the motor aptitudes of young children of approximately the same mental development?

While norms have been set up as indices of general physiological maturity and motor development for young children, and while assumptions have been made concerning the relationship between motor responses and mental abilities of young children, the consistency of the relationship between motor abilities, as evidenced in tests which measure many different kinds of motor reactions, and general mental ability has not been adequately determined by correlation.

For the purpose of determining to what extent the mental age of kindergarten children is revealed in tasks involving motor reactions, the writer has selected a fairly large and comprehensive group of motor tests to be correlated with mental age.

With the exception of some tests which involve only exceedingly fundamental physical movements such as walking, running, jumping, kicking, etc., it would be presumptuous to give the name "motor tests" to many of the tests which are at present assumed to be largely of that nature. In a great number of the so-called motor tests such elements as perceptual discrimination of form, size, direction of attention, analysis, selection, etc., enter. To what extent the motor tests used in this experiment include these and other factors the writer does not attempt to ascertain except in a very approximate way. Some of the tests here used have been employed previously by other investigators, and for these it is possible to compare results. The remainder of the tests here used were constructed by the writer.

The tests used in this study are listed below. A detailed description of each test is given in the discussion of the correlations.[1]

TESTS

Series I

Test 1. Picking up and putting fifteen square blocks into a box.
Test 2. Picking fifteen oblong blocks out of a row of thirty blocks, fifteen square and fifteen oblong, and dropping them into a box.
Test 3. Picking fifteen triangular blocks out of a row of forty-five blocks, fifteen square, fifteen oblong and fifteen triangular, and dropping them into a box.

Series II (Modification of the Wallin Peg Boards).

Test 4. Placing eighteen round pegs into eighteen round holes.
Test 5. Placing nine round and nine square pegs into nine round and nine square holes.
Test 6. Placing six round, six square and six triangular blocks into six round, six square and six triangular holes.

Series III (Montessori Cylinders).

Test 7. Placing cylinders varying in both diameter and height into the proper holes.
Test 8. Placing cylinders varying in diameter only into the proper holes.
Test 9. Placing cylinders varying in height only into the proper holes.

---

[1] Before the directions for the giving of the tests were drawn up, the writer personally experimented with children outside of the experimental group in order to determine comprehension of directions, timing of the tests, etc.

Series IV (Montessori Towers).
Test 10. Placing ten square blocks of equal size, one on top of the other.
Test 11. Placing ten square blocks of graduated sizes in order, one on top of the other.

Series V (Cancellation Tests).
Test 12. Crossing out of circles.
Test 13. Selecting and crossing circles out of a group of different geometrical figures.

Series VI (Modification of Wallin Peg Boards).
Test 14. Same as Test 4, Series II.
Test 15. Placing of colored pegs into the correctly colored holes

Supplementary Motor Tests.
Test 16. Walking the board.
Test 17. Hitting feet together.
Test 18. Turning the wheel.

The data for this study were collected during the fall of 1926 over a period of six weeks. The experimental subjects were the sixty-six kindergarten children in the Training School of Colorado State Teachers College at Greeley, ranging in ages from four years three months, to six years one month. All of these children were given the Stanford Revision of the Binet Scale.

Each of the tests listed above was administered by superior students selected for this purpose from the psychology department. These students were trained personally by the writer until they had gained sufficient facility to administer the tests efficiently. It would no doubt have been more reliable for one examiner to have administered all the tests, but since the study called for each of seventeen different tests to be administered three times to each of the sixty-six children, the total of 3366 individual testings could not have been accomplished by one examiner in the six weeks alloted for collecting the data. Moreover, since the growth and development of children is so rapid at this age, it would have been unreliable to have continued the testing over an extended period of time.

Of the sixty-six children used as subjects in this investigation, only forty-six remained, twenty-three boys and twenty-three girls, upon whom complete records were obtainable. The reduction in the number of children from sixty-six to forty-six was due in large part to absences and withdrawals from the kindergarten. It is believed, however, that the remaining forty-six subjects represent an unselected group. According to Thompson and Pintner, "one criterion for the unselected nature of any group is that the coefficient of correlation of

age and mental age in such a group should be approximately equal to the ratio of the coefficients of variability of chronological and mental ages."[1] The coefficient of correlation for this group between age and mental age is .65, and the ratio of the coefficient of variability of age and mental age is .57. Since these figures are approximately equal, the group may be considered as representative and unselected.

Since, as has been stated, most of our so-called motor tests are not tests of motor ability only, the writer selected a certain number of motor tests which, in his judgment, could be reduced to a simple task involving as nearly as possible only motor elements. This simple test was then considered as the basic one of a series. The other tests in the series were built upon this basic test but were made more complicated by the addition of elements other than motor. The tests comprising a series were given neither consecutively nor in one sitting. Three to four unrelated tests were given at a sitting. This precaution was taken to avoid possible transfer.

As will be noted, some of the tests do not belong to a series, but were considered individually. These are in general almost purely motor in character.

If the children in this group could have been trained in each task until they had reached a stable performance as an index of their maximum efficiency at the time, the score thus obtained would have revealed the child's actual ability. Since it was impossible to train such a large group of children as were included in this study until each had reached his maximum efficiency in each of the seventeen different tasks, only three measures were obtained on each child in each task. Where the task involved the amount done in a certain length of time, the best score of the three trials was used as a measure of the child's ability; where a complete performance was called for, the best time in seconds was used as the score.

Each of the tests used in this experiment was first correlated with mental age and chronological age separately. These correlations, however, do not indicate how much of the ability to do the task is due to the factor of intelligence, and how much to chronological age. Therefore, partial correlations were obtained to determine the extent to which CA with MA held constant, and MA with CA held constant correlated with the tests. The table which follows gives the zero

[1] Thompson, Godfrey H. and Pintner, Rudolph: Spurious Correlations and Relationships between Tests. *Journal of Educational Psychology*, Vol. XV, October, 1924.

TABLE I.—ZERO ORDER AND PARTIAL CORRELATIONS OF CA AND MA WITH THE DIFFERENT TESTS

Tests	Mental age			Chronological age		
	Zero order	PE	CA constant	Zero order	PE	MA constant
Series I						
Test 1	.37	.09	.24	.30	.09	.09
Test 2	.34	.09	.17	.34	.09	.18
Test 3	.63	.06	.52	.42	.08	.02
Series II						
Test 4	.26	.09	−.08	.49	.08	.45
Test 5	.36	.09	.08	.39	.08	.22
Test 6	.67	.06	.46	.60	.06	.30
Series III						
Test 7	.56	.07	.27	.62	.06	.40
Test 8	.48	.08	.27	.46	.08	.23
Test 9	.67	.06	.40	.68	.05	−.03
Series IV						
Test 10	.24	.09	.03	.41	.08	.06
Test 11	.53	.05	.20	.66	.06	.49
Series V						
Test 12	.55	.07	.29	.57	.07	.34
Test 13	.27	.09	.19	.20	.10	.02
Series VI						
Test 14	.26	.09	−.08	.49	.08	.45
Test 15	.35	.09	.16	.36	.09	.19
Supplementary Tests						
Test 16	.47	.08	.15	.58	.07	.41
Test 17[1]	.20	.10	−.08	.40	.08	.36
Test 18	.36	.09	.11	.45	.08	.31

[1] Same as Test 4, Series II.

order and partial correlations between each test and CA and MA. In the discussion of results, whenever correlations between MA or CA and the tests are mentioned, the reference is to the partial correlations.

DESCRIPTION OF TESTS AND DISCUSSION OF CORRELATIONS

Following is a description of the tests used in Series I:

Test 1.—The material for this test consists of fifteen uncolored two-inch cube blocks, and a box into which the blocks can be dropped. The blocks are placed on the floor in a row with intervals of six inches between them. The box is placed at the head of the row at a distance

of three feet from the first block. The purpose of the test is to determine the length of time required for the child to pick up all the blocks, one at a time, and drop them into the box. The child is given three trials and the time in minutes and seconds is recorded for each trial.

Test 2.—The material for this test consists of fifteen two-inch cube blocks, fifteen blocks one-half the size of the two-inch cubes (referred to here as the oblong blocks), two demonstration blocks, and a box into which the blocks can be dropped. According to an irregular arrangement, the smaller and larger blocks are placed in a row on the floor at intervals of three inches between them. The box is placed at the head of the row three feet from the first block. The purpose of the test is to determine the length of time required for the child to select the oblong blocks, one at a time, and drop them into the box. The child is given three trials, and the time in minutes and seconds is recorded for each trial.

Test 3.—The material for this test consists of fifteen two-inch cube blocks, fifteen blocks one-half the size of the two-inch cubes, fifteen triangular blocks, three demonstration blocks, and a box into which the blocks can be dropped. The blocks are placed according to an irregular pattern in a row, two inches between each block. The box is placed at the head of the row, three feet from the first block. The purpose of the test is to determine the length of time required for the child to select the triangular blocks, one at a time, and drop them into the box. The child is given three trials and the time in minutes and seconds is recorded for each trial.

Referring to the partial correlations between the above tests and MA and CA given in Table I, we find that while Test 2 adds seemingly an element of difficulty over Test 1 in asking the child to discriminate between blocks of two different shapes, the correlation with MA, contrary to expectation, does not increase. There are no experimental data here to explain this, but it has occurred to the writer from his observation of the testing, that both Tests 1 and 2 are very elementary for the average kindergarten child, and that the added difficulty in Test 2 is not great enough to distinguish mental abilities at that age.

Referring again to the table, however, we find that the correlation with MA increases significantly for Test 3. It appears that when the child was asked to differentiate between two different kinds of blocks, very little change in difficulty was encountered, but when a third form was included, the perception and discrimination involved

required the child to use more of his mental ability and better differentiated the children according to the mental capacity they possessed. Different results might have been obtained had the child been asked to select the square blocks in all three situations. It is quite evident, however, that CA alone does not, at this age at least, account for differences which the subjects manifested on Test 3. It rather appears that when the difficulty of the test is definitely increased by form perception and discrimination that MA rather than CA accounts for the degree of success or failure.

Following is a description of the tests used in Series II:

Test 4.—Three peg boards similar to the Wallin Peg Board A are used for this test. Each peg board contains six round holes into which round pegs are to be inserted. The eighteen pegs for the three boards are placed in a tray at the right end of the boards. The three boards with their long sides adjacent are placed in front of the child. The purpose of the test is to determine the length of time required for the child to place the eighteen round pegs in the eighteen round holes. Three trials are given and the time is recorded in minutes and seconds for each trial.

Test 5.—Three peg boards similar to the Wallin Peg Board C are used for this test. Each peg board contains three round and three square holes into which round and square pegs are to be inserted. The eighteen pegs, nine round and nine square, are placed in a tray at the right of the boards. The three boards with their long sides adjacent are placed in front of the child. The purpose of the test is to determine the length of time required for the child to place the eighteen pegs correctly. Three trials are given and the time in minutes and seconds is recorded for each trial.

Test 6.—Three peg boards similar to the Wallin Peg Board D are used for this test. Each peg board contains two round, two square, and two triangular holes into which appropriate pegs are to be inserted. The eighteen pegs, six round, six square, and six triangular are placed in a tray at the right of the boards. The three boards with their long sides adjacent are placed in front of the child. The purpose of the test is to determine the length of time required for the child to place the eighteen pegs correctly.[1]

[1] Baldwin and Stecher have employed the Wallin Peg Boards A, B, C and D as offering "one means of approaching the development of motor coordination and form perception." (Baldwin, Bird T. and Stecher, Lorle I.: "Psychology of Pre-school Child," 1924, pp. 74–80.) The procedure employed by the writer

In Series II the writer has again attempted to present the subject with a simple task and then to increase the difficulty of this simple task by adding elements of form perception, discrimination, selection, etc. This series is different from Series I in that it requires a different, and in the writer's opinion a finer type of motor coordination, but similar with respect to the means by which the difficulty of the task is increased, namely, through form perception and discrimination. The effect of adding these elements is noted in the increase in correlations of the second and third tests over the first in this series with MA. The corresponding decrease in the correlations between CA and the second and third tests over the first in this series is an indication that CA alone has less influence on the degree of success than has MA alone. One can see that these tests which Baldwin and Stecher used as measures of motor coordination and control are more than that. Significant is the fact that Board D alone, as used in this investigation, correlates to the extent of .46 with mental age.[1]

Following is a description of the tests used in Series III:

Test 7.—(Board B of the Montessori Cylinders.)

Test 8.—(Board A of the Montessori Cylinders.)

Test 9.—(Board C of the Montessori Cylinders.)

The materials for these three tests are the well-known Montessori Cylinders. Board B consists of a frame into which ten cylinders

---

differs somewhat from that of Baldwin and Stecher. They used but one board A containing six round holes into which round pegs were to be inserted by pupils at ages two, three, four, five and six to determine the number of seconds required for correct completion and also the number of correct placements at these ages. The same procedure was employed by them in the use of Boards B, C, and D. In the present investigation Board B was not employed at all. Moreover, instead of using a single Board A, three Board A's were employed placed side by side and the time required for placing eighteen rather than six round pegs into eighteen round holes was obtained. The same procedure was used for boards C and D. This was done primarily because at the kindergarten age a board containing only six holes would not present a task sustaining enough to differentiate the abilities of the subjects. All the subjects were able to complete the task as it was, but with varying times. Board A was used as the basic test in the series to which elements of form perception could be added by the use of Boards C and D.

[1] The results of this investigation do not agree entirely with Baldwin and Stecher with respect to the comparative difficulty of Peg Boards A, C, and D. There is agreement that Board D is the most difficult as judged by the time for completion. Baldwin and Stecher place Board A as next in difficulty, with B and C as least difficult. The zero correlations with MA in the present investigation would place Board A as least difficult.

varying in both diameter and height are to be fitted. In Board A the cylinders vary in diameter only. In Board C the cylinders vary in height only. These boards were placed in the series according to their difficulty as ascertained by Baldwin and Stecher[1] who found Board B to be somewhat easier than Board A. It was accordingly placed first in this series. The purpose of the tests was to determine the length of time required for the child to place the cylinders correctly. Three trials were given and the time in minutes and seconds was recorded for each trial. [2]

The correlations of MA and CA with these tests are in general consistent with the results obtained in Series I and II. The most difficult test correlates higher with MA than do the simpler ones, and *vice versa* for CA. It is interesting to note in this connection that when CA is held constant Set B is found to be not more difficult than Set A. Board C, correlating to the extent of .40 with MA indicates that this test, in addition to measuring the development of motor control and coordination of the child at different ages in which capacity it was employed by Baldwin and Stecher, also measures to a significant degree the mental development of the kindergarten child. It is interesting to note, in addition, that CA correlates negatively with this test.

Series IV, V and VI are each made up of two, rather than three, tests. Following is a description of the tests used in Series IV:

Test 10.—The material for this test consists of ten square blocks, all of equal dimensions, of the size of the middle block of the Montessori Towers. The blocks are placed in a box within easy reach of the child. The purpose of the test is to determine the length of time required for the child to place the ten blocks, one on top of the other. Three trials are given and the time is recorded in minutes and seconds for each trial.

Test 11.—The materials for this test are known as the Montessori Towers which consist of ten blocks to be placed one on top of the other in order of size from a ten centimeter cube to a one centimeter cube, the purpose being to determine the time required for a perfect

---

[1] Baldwin, Bird T. and Stecher, Lorle I.: "Psychology of the Pre-School Child," 1924, pp. 83–86.

[2] The technique employed by Woolley and Cleveland was followed in the giving of these tests. (Woolley, H. T. and Cleveland, E.: Performance Tests for Three, Four and Five Year Old Children. *Journal of Experimental Psychology*, Vol. VI, 1923, pp. 58–68.)

performance. Three trials are given, and the time is recorded in minutes and seconds for each trial.

Referring to Table I again, we see that for this series the second test shows a decrease in correlation over the first test with both CA and MA. On the basis of the results obtained in Series I, II, and III we had expected that there would be an increase of the correlation with MA and a decrease with CA. Since, however, the correlation decreases for both MA and CA, factors other than the addition of elements of intelligence may be operating. It may be that the Montessori Tower Test is much too difficult for children of this age, and the subjects were thus unable to give a consistent performance in the triple task of discriminating between the blocks, arranging them in order and piling them in order one on top of the other. However, no child failed to complete the test, but the children were extremely variable in their performance. The most difficult test (11) gives a considerably higher correlation with CA than it does with MA. This may be because there is required a greater motor coordination for placing these different sized blocks than the mental ability necessary to direct the task. This task requires considerable motor control. Where the blocks are all of the same size, as in Test 10, the low correlation with mental age may indicate that mental ability is really an almost negligible factor in the successful completion of the task.

Following is a description of the tests used in Series V:

Test 12.—The material for this test consists of a sheet of paper on which appears a checkerboard of circles. The purpose of the test is to determine the number of circles the child can cross out in a given length of time, one minute per trial. Three trials are given. The score is determined from the number of circles crossed out in the given time on each trial.

Test 13.—The material for this test consists of a sheet of paper on which appear outline figures of stars, circles, squares, Maltese crosses, and triangles in chance order. A sample circle is placed at the top of the sheet. The purpose of the test is to determine the number of circles the child can pick out and cross out in one minute. Three trials are given. The score is determined from the number of circles crossed out in the given time for each trial.

The correlations of Test 12 with MA and CA are too low to be significant. They do indicate by comparison with Test 13 that the children are better able to do the simple than the more difficult task. The very nature of the performance required in Test 13 seems to be

too difficult to give increasing scores with either an increase in CA or MA. Why this should be, it is hard to explain as the children seem to be able to discriminate between forms in other types of tasks, especially when they handle the forms. It may be that the newness of the pencil situation has an influence, or it may be that the child does not readily grasp geometrical form when it is given in a single outline only.

Baldwin and Stecher employed the Woodworth and Wells substitution blanks, which are similar to Test 13 in that the child was asked to cross out circles from a number of geometrical forms. By giving the tests to two, three, four, five and six year olds, they found an increase in score from 3.3 at two years to 24.2 at six years.[1] They do not give correlations, however, and how much of this increase in score is due to chronological age and how much to mental maturity they do not suggest. In the narrow range of months between the ages of the kindergarten children in the present experiment, the correlations seem to indicate that CA has very little to do with success.

Following is a description of the tests used in Series VI:

Test 14.—(Same as Test 4, Series II.)

Test 15.—Three peg boards similar to the Wallin Peg Board A are used. Each peg board contains six round holes into which round pegs are to be inserted. Three of the holes in each board are encircled with red. Nine of the eighteen pegs used for the three boards are red. The purpose of the test is to determine the length of time required for the child to correctly place the eighteen pegs, the colored pegs in the colored holes and the uncolored pegs in the uncolored holes. Three trials are given and the time required in minutes and seconds is recorded for each trial.

By comparing the correlations of this series with those of Series II, it is seen that the test involving color discrimination correlates higher with MA than does the one in which the child differentiates round from square pegs. Whether this has any significance or not, the writer cannot say, but the results would suggest that color perception is more closely related to MA than form perception.

In considering together the results obtained from Series IV, V, and VI, and also of the first three series, it would appear that the simple motor tasks correlate higher with CA than with MA. When

---

[1] Baldwin, Bird T. and Stecher, Lorle I.: "Psychology of the Pre-School Child," 1924, pp. 88–90.

the directions are simple and only motor facility is required the correlations with intelligence are relatively low.

As will be noted, Tests 16, 17, and 18 do not belong to a series, but are supplementary tests which are as purely motor in nature as the writer was able to make them. They give some additional correlations between general motor ability and mental and chronological age. Following is a description and discussion of these tests.

Test 16.—The material for this test consists of a "two by four," ten feet in length, supported on its narrow edge by braces. Along the surface of the vertical side are a series of divisions numbered from one to twenty for the purpose of recording the distance the child is able to walk the board before stepping off. Three trials are given and records kept of the distance the child walks on each trial.

This test was designed primarily to measure the child's ability to balance his body. The partial correlations of .41 with CA and .15 with MA show that this balancing ability correlates much higher with CA than with MA.

Test 17.—In this test the child is seated on a chair in a comfortable position so that the knees are just over the edge of the chair to enable the legs to hang freely. The purpose of the test is to determine the number of times the child is able to hit his feet together in fifteen seconds. Three trials are given and records made for each trial.

In this test speed of movement and control of the lower limbs are measured. The correlations of .36 with CA and −.08 with MA show that this type of motor control is much more dependent on CA than on MA.

Test 18.—The material for this test consists of a wooden wheel one foot in diameter mounted in such a way that when the child turns the wheel by means of an attached handle, a noise is made at every revolution, making it possible to count the number of revolutions made. The purpose of the test is to determine the number of revolutions of the wheel the child can make in fifteen seconds.

The variability of performance on this test was considerably less than that of some of the other tests. This probably is due to the simplicity of the task which made it possible for the child to more nearly display his maximum ability. The correlations of .31 with CA and .11 with MA are consistent with the findings from Tests 16 and 17.

In summary, the findings from Tests 16, 17 and 18 support the general findings of the tests in the six series, namely, that when the task is of an almost purely motor character, the correlation is much

higher with chronological than with mental age, and moreover that when the task is almost purely motor, the correlation with mental age is so small as to be almost negligible.

Conclusions

1. It is apparent from the experimental data which have been presented here that motor development alone does not correlate in any practically significant degree with mental age when the subjects are children of the kindergarten stage. However, the coefficients, though low are positive, which fact is in accord with the evidence accumulated during the past quarter century to the effect that desirable traits of all kinds tend to be related, and *vice versa*. The evidence here indicates, however, that motor dexterity alone has little practical value for the prediction of the mental age of kindergarten children.

2. Although seemingly the only means of obtaining a check on the mental potentiality of the infant and very young child is through a study and measurement of his motor reactions, by the time of the kindergarten period motor dexterity alone cannot be taken as an adequate or even approximate guide to the mental development of the child. By the time of the kindergarten period qualities of abstract intelligence, such as concentration, discrimination, etc., are more important than motor development in indicating the mental maturity of the child.

3. While no single test in this experiment has correlated highly enough with mental age to be of utility in the prediction of the individual's mental status for practical purposes, a study of Table I shows that a number of the tests to which elements of abstract intelligence have been added to furnish significant correlations and suggests that a battery of tests involving motor tasks might be constructed which would adequately measure the intelligence of children at the kindergarten age.

CLINICAL OBSERVATION OF EYE MOVEMENTS IN THE RATING OF READING ABILITY

W. R. MILES AND DAVID SEGEL

Psychology Laboratories, Stanford University

Introduction

A discourse by one individual we sometimes designate as a thread or string of conversation. The words, by necessity, have to be in a time sequence. They are strung together on this thread of time and the listener must deal with them accordingly. Such words have but little spacial connotation; they are not thought of as proceeding from left to right or from top to bottom. Auditory impressions and interpretations of these depend primarily on this temporal arrangement. The motor pattern of the listeners may be quite diverse and still permit of adequate listening. Writing words to be presented to the eye may also be thought of as a string of symbols. Here the placement or the spacial arrangement rather than time sequence is significant

A spacial string of symbols obviously may be spread out on a wall or sheet of paper in many different ways without having the lines cross each other. For example, it could be stretched out in a straight line, wound up in a flat spiral (circular or square), arranged in a short zig-zag course and so on. All of these are about equally logical and intelligible if you know what method has been followed. The scheme of cutting up the string into short pieces and placing these in a column arranged from top to bottom with each piece to be gone over in the order of left to right is only one of many possible arrangements. It happens to be the way that a large portion of the earth's inhabitants has hit upon for placing written language in straight lines on the page. English and most European languages have this arrangement. Semitic languages and Chinese have used other dispositions. In any language before you read you must have a notion of the *way* to read, that is, of the reading direction.

The first principle of reading that the child must learn is where to go next. It is a game of finding your way, somewhat like progress through a fairly familiar maze path. Even as yet teachers of reading seem to have no very clear cut method of communicating this first principle. It ought to be displayed very thoroughly for the reason that at first the child can see no more sense in playing the reading game one way than another. He has of course to go around the school

house or around the fence on the playground but his eye can go anywhere in the printed picture or on the printed page; there are no natural eye-barriers that route him along one course; these he must create for himself as he understands the scheme of the material. The child may discover it for himself but probably can profit by being directly told. In this game that his eyes play with the book, the eyes start at the top, walk across slowly on the narrow path to the right, jump clear back, step down to the next path, walk across slowly, jump back, step down and so on. If there is a picture on the page he can stop the regular walking and jumping and make a jump to the middle of the picture from which point he *runs around* all over it even into its corners.

This pattern of motor behavior can be easily communicated to the individual child, for example, by making use of the song book and a song which the child knows fairly well. While the words are being sung, the teacher moves her finger along the line, not steadily but as if taking steps with the finger thus making natural fixation points or stops for the eye. Then she quickly jerks the finger back to the left to begin the new line. To many teachers this motor pattern seems too simple to trouble about. They apparently think that it will automatically adjust itself. However, to attain the motor control necessary for smooth reading probably involves the mastery of a more complex motor coordination coupled with the understanding of directions of where to go next than is required for playing hopscotch. The position of a child at the desk and the manner in which he holds the pencil for writing are features of motor behavior which we can easily observe and in large measure directly correct. Through pronunciation we have a fairly good check on tongue movements and can easily drill the child repeatedly until a certain sequence and temporal relationship is achieved for purposes of articulation. Eye movements are peculiarly hidden. The person himself is scarcely conscious that they are taking place as specific items. They do not declare themselves to the teacher by any noisy clatter. The movements are quick and small and moreover between the eye and the object there is a big gulf fixed which makes it difficult for the teacher to know exactly at what point the child's eyes are directed.

Investigations of the reading process have generally consisted in (1) a fairly elaborate observation of the motor or physical side including eye movements, head movements and vocalization or, (2) by means of reading tests, measuring the product of reading ability with little

or no attention on the motor coordination side. The two methods have been carried on almost independent one of the other and the amount of correlation of the two sorts of results has been meagre. Both methods would appear to be of value as they are complementary one to the other and each can yield results of importance particularly in reference to the reading of young children. The test method is at present in the ascendancy so far as judged by number of studies and extent of use. A condition that has favored this popularity of the reading test has been the difficulty and general instrumental complexity necessary for making the eye movement investigations.

Photographic registration of eye movements in reference to different grades of reading ability and different kinds of reading material should, of course, not be neglected. However there appear to be some very simple clinical methods that can be easily used in the school room and that can be depended upon to yield significant results for diagnostic and remedial work.

As is well known, Erdmann and Dodge made their original investigation without recording eye movements. They observe by the use of mirrors and also by the use of a small telescope. Freeman[1] has devised a mirror arrangement which has apparently given good results and recently Newhall[2] has described a simple equipment. This latter includes light, head rest, copy holder and lens for magnification of the movements to aid in directly observing them. We have chosen an even simpler scheme and one which perhaps is more in keeping with our fundamental habits in observing the eyes of those with whom we talk. This has been called the "peep-hole method."[3] It is based on the principle generally employed in optical instruments and eye examinations, that is, of observing through a small hole in the middle of a mirror or of a dark screen. The observation is thus from what would naturally be the primary line of fixation. As we face an individual we can accurately judge the position of his line of sight. We know whether he is looking us straight in the eye or we know about how far to one side he is looking. Thus if we cut a small hole in the center of a page of reading material, hold the page in front of the reader's face, and place one of our eyes behind the opening we

---

[1] Freeman, F. N.: "Experimental Education." Boston, 1916, pp. 95–139.

[2] Newhall, S. M.: Instrument for Observing Ocular Movements. *American Journal of Psychology*, Vol. XL, 1928, pp. 628–629.

[3] Miles, W.: The Peep-hole Method for Observing Eye Movements in Reading. *Journal of Genetic Psychology*, Vol. I, 1928, pp. 373–374.

have the best possible viewpoint from which to see the succession of eye movements and fixations, both as to the placement of these and for counting them. In addition the observer's face is hidden behind the paper reducing distraction to a minimum. The reader's head is unrestrained and his mouth is in full view so that head movements and lip movements may be observed.

Procedure

We chose pupils from Grade III as subjects for our investigation. At that level there is a fairly wide range of reading ability, but as a rule the speed of reading is not too fast for the type of observation planned. We may suppose also that young children are more likely than adults to show irregularities from the typical motor pattern. For testing the product of reading we used the Gates Primary Reading Tests and the Gates Silent Reading Tests. The former presumably measure to some extent the primary adjustments or mechanics of reading. The norms are for grades from 1 to 3.6. The Gates Silent Reading Tests concern reading at a more mature level. These tests are made primarily for Grades III to VIII inclusive. The Primary Reading Test has three parts and the Silent Reading Test has four parts. The average educational age found for each set of tests was used in our investigation. Independently of our tests and clinical measurements, we had each pupil rated by the teacher. A reading ability scale of from one to five was used. These ratings constituted an additional "measure" of reading ability and were kept secret from the investigator until all the tests were finished.

We took account of the following factors as external evidence of reading:

1. Speed of reading.
2. Fixations.
3. Regressives.
4. Confusions (rather detailed examination of a word).
5. Head movement (turning of the head to follow line).
6. Vocalization.

We considered that the best reading performance was rapid, with few rather than many fixations, with few or no regressives, with few or no confusion periods, with little or no head movement and with little or no vocalization. The peep-hole[1] technique was found adapted

[1] Miles: *Op. cit.*

to observing fixations, regressives, confusions, head movements and vocalization all at once. Speed was recorded by stop-watch. From our experience we judge that children can be thus observed without interfering with their normal reading habits.

For our work we chose two reading selections[1]—one which was adapted normally to Grade II children—and the other to Grade III children. These selections were mounted on 9 × 12 inch cards and the peep-hole $\frac{3}{16}$ inch square was cut in the center. The card used with Grade III children is here reproduced about one-third size (see Fig. 1).

Each of the fifty-nine Grade III pupils examined was asked to read these selections. Observing through the peep-hole the number of fixations, regressives, and confusions were recorded as the child read by tallying on a pad of paper as follows:

| | | − | | | | 0 | | − | | | | | | | − |

The illustration given is the tallying for three lines of reading. The | indicates a fixation, − a regressive, and 0 a confusion. This recording, which became somewhat automatic, allowed the observer to judge at the same time the amount of head movement and vocalization. We were thus able to get a composite score for each subject on each of the two selections. The scores for fixations, regressives, and confusions are objective, in part, in that they were records made by the observer at the time. It is true that these observations are subject to systematic and other errors. In so far as the errors are systematic, the results would be affected only as compared with those of other investigators. The average number of fixations, regressives, and confusions per line was considered to be the score. The head movement was rated on a five point scale ranging from no movement to almost a complete substitution of head movement for eye movement. Vocalization was also rated on a five point basis on the evidence presented by lip movement and sound production.

The observation and necessary recording took approximately five to six minutes for each child. The investigator knew nothing whatever of the ability of any child before the observation took

---

[1] These selections were Starch's Reading Test Series A, No. 1 and No. 2 (see Starch: "Educational Measurements," New York, 1916, pp. 22, 23). Originally designed for Grades I and II but found adapted for Grades II and III in California schools.

place, and the teachers in making their ratings had no knowledge of the results of the observations or of the results on the Gates tests.

The scores made on the factors of external evidence of reading ability were changed into scores of equivalent range 0–8. The first three factors, time, fixations, and regressives, were weighted twice as much as the others. Upon investigation this was found to be about as accurate weighting as we could get. Then a composite score was

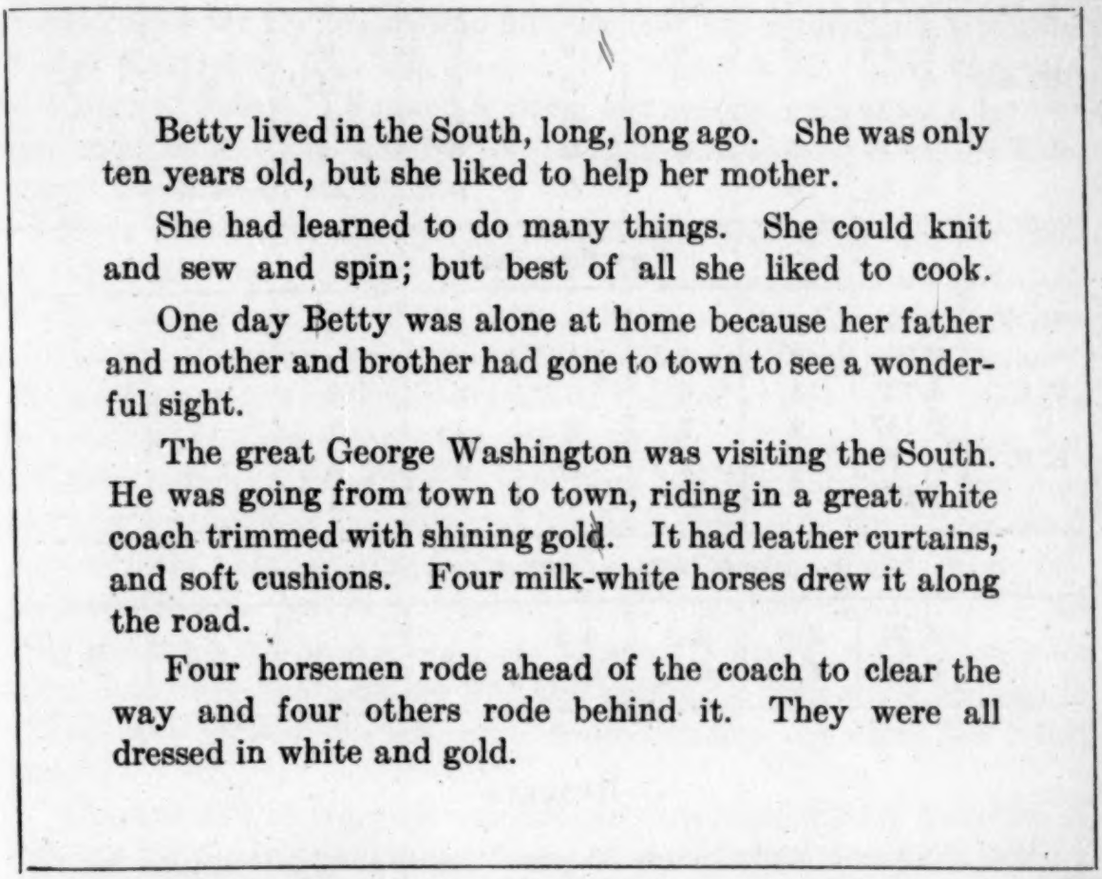

Betty lived in the South, long, long ago. She was only ten years old, but she liked to help her mother.
She had learned to do many things. She could knit and sew and spin; but best of all she liked to cook.
One day Betty was alone at home because her father and mother and brother had gone to town to see a wonderful sight.
The great George Washington was visiting the South. He was going from town to town, riding in a great white coach trimmed with shining gold. It had leather curtains, and soft cushions. Four milk-white horses drew it along the road.
Four horsemen rode ahead of the coach to clear the way and four others rode behind it. They were all dressed in white and gold.

FIG. 1.—Reading Selection B for observing clinically the mechanics of reading. Observing is done through the opening in the middle of the figure. The broken lines in the selection assist the examiner in following definitely the progress of the child over the page.

obtained from these transmuted weighted scores. Our composite score was obtained by multiplying the time score by two, the fixation score by two, the regressive score by two, and adding in the confusion score, the head movement score and the vocalization score. Correlations were calculated between these composite scores and the results on the Gates tests and for the teachers' ratings. Also correlations were computed between the teachers' ratings and the Gates tests.

TABLE I.—READING RESULTS FOR GRADE III PUPILS

Subject	Time	Fixations	Regressives	Confusions	Head movements	Vocalization	Reading age	Teacher's rating
Very Good Readers								
L. M.	A 22	3	0	0	0	0	14.0	1
	B 23	3	0	0	0	0		
B. O.	A 18	3	0	0	0	0	11.1	1
	B 24	3	0	0	0	0		
H. Ki.	A 15	3	0	0	0	0	9.7	2
	B 24	3	0	0	0	0		
S. F.	A 18	3	0	0	0	0		
	B 21	3	0	0	0	0	9.6	1
Very Poor Readers								
M. Y.	A 240	6	4	4	4	4	8.1	5
	B 360	7	5	5	4	4		
P. G.	A 125	8	6	6	4	4	8.4	4
	B 187	9	7	7	4	4		
E. S.	A 72	6	4	4	4	4	8.4	5
	B 108	7	5	5	4	4		
Average for Fifty-nine Third Grade Students								
	A 56	3.8	0.2	0.2	.9	1.1	9.3	2.9
	B 82	4.7	0.8	0.4				

RESULTS

We have not undertaken to give the individual results for the fifty-nine Grade III pupils who were measured. Table I shows what was found for the four best readers and for the three poorest readers and gives as well the averages for the entire group. In considering the reading ability range exhibited for these pupils it is found that the selections could be read in from fifteen to twenty-five seconds by the best readers, while the poorest ones required from one hundred to three hundred seconds; that is, ten times as long to go over the same material. The number of progressive eye fixations per line did not vary in this same ratio. The best readers used an average of three fixations per line, the poorest ones used six to nine. Any such ratio in the number of eye movements is complicated by the fact that the

poor readers made on the average several regressive movements, four to seven per line and exhibited as well confusion intervals. The poor readers also used head movements to a marked degree in substitution for eye movements and tended to vocalize or to read the material in a loud whisper. The four best readers as well as several others in the first quartile of the group showed practically no regressive movements or confusion intervals and made little or no head movement as they read silently without lip movement. The average reading time for the group was fifty-six (SD 40) seconds for selection A and eighty-two (SD 60) seconds for Selection B. This amounts to a reading rate of 1.8 and 1.5 words per second respectively for the two selections. The number of reading pauses per complete line were 3.8 for A and 4.7 for B.

If we count that the eye movement time was about six per cent of the total time required for the reading, *i.e.*, ninety-four per cent of the time being spent in fixations, it is found that the average pause or fixation duration for A was 0.99 second and for B, 1.03 seconds. We see here a further demonstration of the well known law of fixations in reading: the duration of fixation tends to remain the same for different lines and for different selections but the number of fixations per line increases as the material becomes more difficult. The usual results from photographic recording of reading for adults shows the fixation time to be about one-fourth to one-third second. We see that the very good readers in our Grade III group, those that were using but three fixations per line and reading a fifteen line passage in fifteen to eighteen seconds were approximately equalling the adult rate.[1]

The number of regressive movements averages 0.2 for Selection A and 0.8 for B; and the values for confusion intervals are 0.2 and 0.4 respectively. In studies by photographic technique we do not obtain head movements or lip movements. The head and mouth rest are intended to eliminate these. However, it is perfectly obvious from our clinical results, and from what every teacher knows, that the poor readers tend to exhibit pronounced head movements and to read aloud or whispering. These symptoms are flagrant signs telling that either the material is much too difficult or that individual instruction of a remedial nature should be undertaken.

[1] The results cited by O'Brien: "Silent Reading," New York, 1922, pp. 236–247 for Grade IV readers, agree approximately with our results on the best Grade III subjects.

The reading selections employed in our clinical examination were very easy—too easy for the best readers. From Table I we see that the habit of using three pauses per line sufficed for both passages of material for them. Selection A contained one hundred two words and selection B, one hundred thirty-two. L. M. and S. F. covered more words per second when reading the latter, that is Passage B. B. O. covered them both at the same rate, but H. Ki. was slower on B. The situation is quite different for the poor readers. Here in each case the pupil was much slower on the B passage. At the reading rate used for A they should have covered in their various time scores for B, one hundred fifty-three words whereas the number was one hundred thirty-two. For these pupils the number of fixations, the number of regressive movements and of confusion intervals is greater for selection B. Head movements and lip movements show no difference for the reason that these features are very prominent parts of the motor pattern for these pupils. When we correlate the time scores shown by the fifty-nine pupils on the two passages we find by the product-moments method the result to be $+ .90 \pm .02$. The reading rate for a given Grade III pupil is therefore found to be fairly characteristic for him when he tries different selections. When the time scores were arranged in the form of a scatter diagram for working out the correlation it was found that eight of the pupils constituted an outstanding group of poor readers, well separated from the scatter representing the other fifty-one pupils. The two poorest of these were ranked five in the teacher's ratings and the other seven all received the rank of four. There were, however, four other pupils in the entire group who received the rank of five so that slowness of reading did not wholly determine the teacher's estimate.

The reading age determined by the Gates Silent Reading Tests is indicated in the second column from the right. The three fastest readers did not score highest in this particular. The slowest reader (although he did not use the largest number of fixations or make the largest number of regressives) did score lowest in the Gates Test. The scores for the whole group range from 14.0 to 8.1. The average reading age rating was 9.3 years. This corresponds to a grade placement of 3.6 while the actual grade placement was about 3.4. The teachers ratings at the right of the table correspond fairly well with the clinical results for observed eye movements and with the test results. Some of the correlations found by the product-moments method, uncorrected for broad catagories, were as follows:

Clinical score with Gates Silent Reading Tests (59 cases).....	.66 ± .05
Clinical score with Gates Primary Reading Tests (31 cases).....	.75 ± .05
Clinical score with teachers' ratings (59 cases).....	.70 ± .05
Teachers' ratings with Gates Primary Reading Tests (31 cases)..	.73 ± .06
Teachers' ratings with Gates Silent Reading Tests (59 cases).....	.84 ± .03

Conclusions

1. Photographic registration of eye movements is not to be neglected in studying the reading process and the performance of given individuals. However, a simple clinical method may serve in the school room for selecting cases to study or for undertaking immediate remedial training.

2. The clinical method adopted, termed the "peep-hole method," consists in *observing from the center of the page which is read.* The examiner places his eye behind a small opening at the middle of the page of material and holds the page in front of the pupil. He can therefore directly see the position of the line of sight, fixations, head movements and lip movements.

3. A group of fifty-nine unselected Grade III students were examined. The best of these readers show fixation durations which correspond closely to findings for adults. They make practically no regressive movements, show hardly any confusion intervals, make little or no head movement or lip movement.

4. The very poor Grade III readers require ten-fold as much time to cover the same material, use twice as many fixations, make many regressive fixations and conspicuously show head movements and lip movements.

5. A correlation of about .70 ± .05 was found between the clinical results and results gotten with the Gates Reading Tests and by teachers' ratings. The clinical results have the advantage over the test and teachers' rating results, in being more analytical in that they show the motor habit side of the difficulty which we call poor reading ability. From such clinical observation we can easily determine whether the child thoroughly knows the *way* to read.

SOME SCIENTIFIC APPROACHES TOWARD VOCATIONAL GUIDANCE[1]

CHAS. E. LIMP

Department of Public Instruction, Madison, Wisconsin

From the earliest primitive developments of the human race, man has had vocations. As civilization advanced, the needs and wants of the peoples increased until at present our complex civilization creates vocations for its thousands of needs. For a long period in the early development, the son took up the vocation or trade of the father—a custom still prevalent in many countries today. In more recent years, man has broken away from this custom. He seeks to do the things that appeal to him the most and the thing for which he is best fitted. Ever since there has been a vocation there has been some form of guidance.

The most potent factors in determining success and efficiency in any vocation are interest, ability to do, and perseverance. The trial and error method is not only expensive in time and effort but wasteful in terms of efficiency. If some device or battery of tests could be worked out whereby one's aptitudes or abilities could be determined, a great stride forward would be made in reducing the waste and efficiency in selecting vocations.

In attacking a problem on a scientific basis, data of some kind must be collected or observed and evaluated. The following steps are suggestive:

1. See the problem.
2. Formulate the problem.
3. Secure necessary data by tests and experiments.
4. Arrange and classify findings.
5. Formulate hypothesis.
6. Verify hypothesis by repeated experiments.
7. Suspend judgments and state findings in quantitative terms, not in opinions.

For the past two decades tests have played an important rôle in education. We have had educational tests in abundance covering almost every phase of subject-matter in the curriculum from kindergarten to the university. Intelligence tests have, perhaps, played even a more important part.

[1] This is the fourth and last of a series of articles in aptitude testing.

From the discussions of leading educators at the Department of Superintendence, Boston, one would gather that environment has a marked bearing upon the supposed fixity of the IQ. The American people have run amuck over tests and intricacies of rows of figures. The whole testing program has become quantitative instead of qualitative. The testing movement has reached a position in education comparable to what Thorndike calls a plateau in the learning process, and until the bonds of the testing program are sufficiently strenghtened to reach the higher level, we are marking time. Yet, tests as instruments of measurement with their great contribution to education and their future possibilities, when computed in terms of efficiency, are only a crude instrument and far from the goal of what might be attained.

Not long ago someone predicted that the greatest value of tests would not be that of testing ability or achievement, but that the emphasis would change to that of educational aims and ideals; that we shall be more concerned than we now are about the significance of intelligence—both general intelligence and special abilities—for educational and vocational guidance.

Early in 1922 the writer began an investigation in order to determine for ninth year students whether ability to do shorthand and typewriting could be predicted. The results have already been published.[1] Simultaneously two other investigations that need to be mentioned here are being carried on in the field of vocational guidance.

The Study at the University of Minnesota

Dean J. B. Johnston[2] has done some very interesting work in predicting success or failure in college at the time of entrance. He has shown rather conclusively that ability to do or ability not to do the work and the grade or quality of work done can be predicted with a certain degree of precision and accuracy in most cases. During the

---

[1] Hull, Clark L. and Limp, Charles E.: The Differentiation of the Aptitudes of an Individual by Means of Test Batteries. *Journal of Educational Psychology*, Vol. XVI, Feb., 1925, pp. 73–88.

[2] Johnston, J. B.: Predicting Success in College at the Time of Entrance. *School and Society*, Vol. XXIII, January 16, 1926, pp. 82–88.

Johnston, J. B.: Predicting Success or Failure in College at the Time of Entrance. *School and Society*, Vol. XIX, June 28, 1924, pp. 772–776; Vol. XX, July 5, 1924, pp. 27–32.

Johnston, J. B.: Tests for Ability before College Entrance. *School and Society*, Vol. XV, April 1, 1922, pp. 345–353.

years 1923–1924–1925, one thousand eighty-eight students were given intelligence tests. It was predicted that two hundred eight of this number would fall below the average grade of C. Of these two hundred eight, only three did secure the C standing. The accuracy of the above prediction is so much greater than that which is secured in choosing employees in the industrial world or in some other forms of prediction in human affairs that the error of less than one and one-half per cent may be regarded as negligible. This means that the injustice done to three persons by advising against college work is wholly incomparable to the injustice done to two hundred eight persons when no action is taken and therefore advice to go to college is implied.

Predicting distinguished scholarship is much more difficult. In 1925–1926 a list of freshmen was prepared with the hope of including all those who would secure an average of B in their studies. Of the whole group thirty-nine did secure the B average, and of these thirty-six were included in the advance list. Such predictions can be made with some degree of certainty.[1]

The Study at the University of Illinois

In the fall of 1923 the Bureau of Educational Research, University of Illinois[2] secured data from about twelve thousand seniors in high schools which gave a rating as to their mental ability. A year later the scholastic records of the same seniors were secured. The records of those who went on to institutions of higher learning and finished the freshmen year were correlated with their scores in mental ability and scholastic attainments while seniors in high school. Coefficients of correlation were obtained as high as .60, but the writer in his summary and conclusions states that even this is little better than poor guesses, and quoting from this summary we read: "Since both for the purpose of determining the admission of applicants to college and of personnel work with students who have admitted, it is desirable to predict the quality of their scholastic work, it is better to make use of such bases of prediction as we have, than to rely upon pure guesses. Therefore the use of high school marks, entrance examination results, or intelligence test scores for this purpose is better than that of no criterion at all."

---

[1] Johnston, J. B.: Predicting College Success for High School Seniors. *Vocational Guidance Bulletin*, 2, No. 2, November, 1927, Minneapolis Public Schools.

[2] Predicting the Scholastic Success of College Freshmen. *University of Illinois Bulletin*, 25, No. 2, September, 1927.

For those who have not read the writer's previous articles[1] describing the experiment, the problem was to find out whether ability to do shorthand and typewriting could be predicted. In the selection of tests, any test that looked at all likely as a possibility in the entire range of possibilities was selected; intelligence tests, motor tests, educational tests, will-temperament tests, etc. In the original experiment one hundred seven students came through with tests and criterion records complete in every detail. The five tests selected for each battery were the following:

TABLE I.—ZERO-ORDER CORRELATIONS AND FOURTH-ORDER PARTIAL CORRELATIONS OF THE TESTS SELECTED

Tests correlating highest with shorthand	Correlation		Tests correlating highest with typewriting	Correlation	
	Zero-order	Fourth-order partial		Zero-order	Fourth-order partial
(1)	(2)	(3)	(1)	(2)	(3)
Logical selection (4)[1]	+.33	+.21	Best answer (3)	−.10	−.16
Sentence meaning (6)	+.06	−.06	Sentence meaning (5)	−.19	−.25
Recognition spelling (17)	+.53	+.41	Recognition spelling (17)	+.49	+.50
Courtis multiplication (39)	+.40	+.19	Courtis multiplication (39)	+.33	+.21
Word meaning (40)	−.50	−.20	Word meaning (40)	−.23	−.16

[1] Tests numbered 3, 4, 6, and 40 are taken from Terman's Group Intelligence Tests. Test No. 17 is taken from Hoke's Stenographic Test. Test No. 39 is taken from Courtis' Arithmetic Test.

In 1923 the tests were repeated in the same school with even better results. In 1924 the tests were given to a larger number of students in another school. Again the tests verified the results of both the original and repeated the experiments. Each year the tests have been given

[1] Hull, Clark L. and Limp, Chas. E.: *Op. cit.*

Limp, Chas. E.: The Use of the Regression Equation in Determining the Aptitudes of an Individual. *Journal of Educational Psychology*, Vol. XVI, Sept., 1925, pp. 414–18.

Limp, Charles E.: A Work in Commercial Prognosis. *Journal of Educational Research*, June, 1927, pp. 48–56.

and every repeated experiment has proven the validity of the original experiment. Two years ago, the tests were given to a group of sixteen students in a class of vocational aptitude testing at the University of Wisconsin. This was a mixed group so far as the teaching profession is concerned. Two were commercial classroom teachers, three were manual arts teachers, one was an art teacher, three were high school principals, etc. The predicted grades for the two commercial teachers were one hundred ten and one hundred twelve respectively, and one machine-shop teacher who stated before taking the tests, that he was unable as a high school student to master shorthand and typewriting, received a predicted score of forty-eight.

The battery of tests for shorthand in the original experiment gave an R of .6116 and the typewriting battery an R of .6347. The error of estimate for the shorthand was 7.2; for typewriting, 5.8. If the error of predicting and forecasting success of a given aptitude by means of a battery of tests is equal to an SD of the criterion, the test is of no value whatsoever. Conversely, the standard error of forecasting a criterion by means of a battery of tests is zero if the tests are perfect. The actual efficiency of any test, or the value of tests is definitely measured by this error of estimate.

The regression equation which gives the weighted test scores for shorthand was found to be

$$X_c = +.681X_2 - .112X_3 + .277X_4 + .562X_5 - .277X_6 + 62.63$$

In this equation the subscript c refers to the predicted criterion and the subscripts, 2, 3, 4, 5, and 6 stand for tests 4, 6, 17, 39, and 40, respectively. In the regression equation for typewriting

$$X_c = -.213X_1 - .355X_3 + .256X_4 + .444X_5 - .162X + 73.33,$$

the subscript c refers to the predicted criterion and the subscripts 1, 3, 4, 5, and 6 stand for tests 3, 6, 17, 39, and 40, respectively. By means of these regression equations, estimates or forecasts, were made in terms of school grades for each pupil in both shorthand and typewriting.

Table II illustrates the use of these regression equations. John Doe made the test scores in shorthand as shown in column 2. (The numbering of the tests in columns 1 and 5 correspond to the subscripts in the regression equations. Thus test No. 2 in Table II is the original test No. 4.). John Doe's score of 2 on the test No. 2, according to the regression equation for shorthand, must be multiplied by + .681 (column 3).

TABLE II.—THE USE OF THE REGRESSION EQUATION IN COMPUTING THE PREDICTION GRADE

Test number	Shorthand score			Test number	Typewriting score		
	Test score	Multiplier or weight	Weighted score		Test score	Multiplier or weight	Weighted score
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
.....	..	.....	.....	1.....	1	−.213	− 1.70
2.....	2	+.681	+ 1.36				
3.....	4	−.112	− .45	3.....	4	−.355	− 1.42
4.....	44	+.277	+12.19	4.....	44	+.256	+11.26
5.....	8	+.562	+ 4.50	5.....	8	+.444	+ 3.55
6.....	4	−.277	− 1.11	6.....	4	−.162	− .65
Constant (*K*)...	..	.....	+62.63	Constant (*K*)	..	.....	+73.33
Total predicted score.....	..	.....	79.1	Total predicted score.....	..	.....	84.4

This gives a product or weighted score of + 1.36 (column 4). When the arithmetic required by the regression equation for shorthand has been completed, John Doe's predicted grade for shorthand is found to be 79.1. A similar computation, using the regression equation for typewriting, results in a predicted grade of 84.4 for that subject.

If a battery of tests can be devised covering the entire range of human aptitudes, each test measuring an individual aptitude, independent and no overlapping of one test with the other, vocational guidance could be reduced to a science. For, from the entire range of individual tests, groups of regression equations for four or five individual tests measuring the particular traits that make for success in each particular vocation, could be worked out. While the battery of tests selected in this investigation for shorthand and typewriting have not been altogether satisfactory, yet the average error of prediction for over one thousand students in various high schools of the state was 6.3 for shorthand and 5.2 for typewriting in terms of school marks, which is much more reliable than mere chance guessing.

Commercial classroom teachers would find these tests very helpful in advising students in the selection of commercial courses, and may well afford to use them. By administering the tests the predicted score would reveal what school mark or grade the student would be likely to recieve if the course were taken.

The tests in each battery have not been assembled and published as a battery. The individual tests, however, may be purchased[1] and assembled. The writer is submitting the regression equation for both shorthand and typewriting and a table for computing the predicted scores (Table II).

The experiment is still being carried on with the hope of improving and refining the regression equation in order that the tests may render a greater service.

What we need in the field of research is not necessarily more investigations but more investigation in the problems already started, and not so many "tentative conclusions" that appear at the end of published research problems. However, until our present instruments for measuring aptitudes are greatly improved, vocational guidance in terms of efficiency will not rise much beyond the level of pure chance guessing and giving advice.

[1] Terman Tests are published by the World Book Company. Hoke's Tests may be secured from the author, Elmer R. Hoke, Annville, Pa. Courtis Arithmetic Tests may be secured from S. A. Courtis, 1807 East Grand Boulevard, Detroit, Michigan.

A NOMOGRAPH FOR ESTIMATING THE RELIABILITY OF A TEST IN ONE RANGE OF TALENT WHEN ITS RELIABILITY IS KNOWN IN ANOTHER RANGE

EDWARD E. CURETON AND JACK W. DUNLAP

Territorial Normal and Training School, Honolulu, Hawaii

Kelley[1] gives the equation connecting the reliabilities and standard deviations of a test in two ranges of talent as

$$\frac{\sigma}{\Sigma} = \frac{\sqrt{1 - R}}{\sqrt{1 - r}}$$

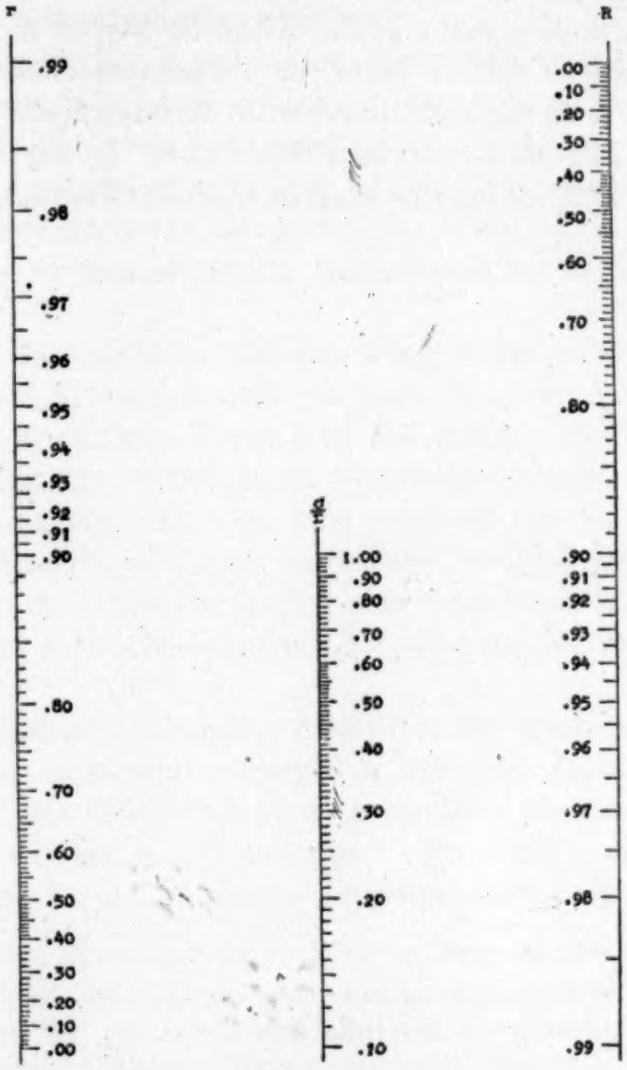


FIG. 1.

[1] Kelley, T. L.: Reliability of Test Scores. *Journal of Educational Psychology*, Vol. XI, No. 5, May, 1921, p. 370.

where

σ is the standard deviation for the shorter range
Σ is the standard deviation for the longer range
r is the reliability coefficient for the shorter range
R is the reliability coefficient for the longer range

Solving for r and R, we have

$$r = 1 - \frac{\Sigma^2(1 - R)}{\sigma^2}, \; R = 1 - \frac{\sigma^2(1 - r)}{\Sigma^2}$$

The nomograph solves both of these equations. It is necessary first to compute the ratio σ/Σ. Then by laying a straightedge (a stretched thread or rubber band, or a line scratched on a piece of celluloid) from R on the right hand scale through the σ/Σ value on the center scale, it will cut the left hand scale at r. By laying the straightedge from r to σ/Σ, R may be read in a similar fashion from the right hand scale.

THE PROGNOSTIC VALUE OF THE LOW RANGE ARMY ALPHA SCORES

ALFRED H. JONES

Department of Psychology, University of California at Los Angeles

Probably one of the greatest problems facing the administrators in higher education is the elimination and advisement prior to matriculation of those students unfit for collegiate work. Such a procedure avoids the alternate handling of the situation, which is usually withdrawal or dismissal. This is not only expensive for all concerned, but it also leaves behind it a sense of failure, which in many cases proves to be an indeterminable handicap.

To meet this situation, freshman orientation weeks, pre-registration advisory schedules, and testing programs have been instituted. The Army Alpha Test has been used to considerable extent in trying to determine any indication of low ability for prognostic use. In testing for the segregation of this group the Army Alpha is particularly adapted because it was originally constructed for similar use in the Army.

In view of this problem the following paper substantiating the use of the Army Alpha for such purposes is given. In September, 1926, prior to registration Form 5 of the Army Alpha was given to six hundred fifty-nine students in Teachers College, University of California at Los Angeles. Of this number the forty who made scores of less than one hundred have been selected for study in this paper. The median for the group was one hundred forty with a probable error of nine, the range seventy-five to one hundred eighty-nine.

The first semester's scholastic records in the university were tabulated in order to determine correlation with the Alpha scores. The academic work was then evaluated in terms of the number of grade points above or below a "C average." For clarity of meaning the following excerpt from University Announcement is given.

SCHOLARSHIP RULES: DISMISSAL FOR FAILURE TO MEET MINIMUM REQUIREMENTS

1. Any student in these colleges whose scholarship in any half-year falls *below* an average grade of "D" per unit of work undertaken, will be dismissed.

2. (*a*) Any student in these colleges whose entire university record, computed to the close of any half-year, shows a cumulative deficit of ten (10) or more grade points (regardless of the number of units passed) will be placed on probation.

(*b*) If the student fails during any single half-year while on probation to maintain an average grade of at least "C" per unit of work undertaken, dismissal

will follow; otherwise, probationary status will be continued until such time as the cumulative grade point deficit be reduced to an amount less than ten (10).

3. By special action of the Committee on Disqualified Students, any student who in any single half-year fails to attain grade points in an amount equal to the number of units undertaken, may be placed on probation for one-half year, and be required during the probationary half-year to maintain at least a "C" average under the penalty of dismissal.

SCHOLASTIC STANDING OF STUDENTS ENTERING FIRST SEMESTER, 1926–1927, WHO HAVE ALPHA SCORES OF LESS THAN ONE HUNDRED, IN ORDER OF MERIT. RECORD COVERS TWO SEMESTERS UNLESS OTHERWISE NOTED

Field of study[1]	Army Alpha score	Grade points: − below C average + above C average	Disposition	
		Plus		
G. E.	94	34		Above C average
Kgn. P.	97	5		
G. E.	97	9	St. Brd. Canad., withdrew, age 30	
L. S. Span.	94	7½		
Kgn. P.	98	4		
Art	90	1		
Jr. H. S.	96	0		
		Minus		
Art	85	½		
G. E.	98	1		
G. E.	76	1½		
G. E.	85	2½	Admitted Adv. Stdg., age 22, first semester only	
G. E.	99	5½		
G. E.	98	5½	Withdrew first semester, record only	
H. E.	75	6		
H. E.	97	6		
G. E.	85	6		
Com.	90	6½		
Kgn. P.	95	6½	Withdrew first semester, record only	
Mech. Arts	95	7		
Music	92	9		
Chem.	60	10	Probation	On probation
Kgn. P.	98	10	Probation	
Kgn. P.	94	11½	Probation	
G. E.	83	12	Probation	
Kgn. P.	97	14	Probation	
Home Ec.	88	14½	Probation	
Kgn. P.	77	15	Probation end first semester, withdrew first mid-term of second semester	
L. S.	96	17	Probation	
G. E.	99	....	Withdrew after second mid-term with D, DD, E	Withdrew during first semester
H. E.	94	....	Withdrew before first mid-term	
Kgn. P.	99	....	Withdrew before first mid-term	
H. E.	85	....	Withdrew first mid-term, 7½, E; 4, F; 3, D	
G. E.	98	....	Withdrew after first mid-term of 3½, E; 9, D; 4, F	
P. E.	88	....	Withdrew after second mid-term of D, DD, E	
Kgn. P.	95	....	Withdrew before first mid-term	
		Minus		
P. E.	94	18	Dismissed—reinstated	Dismissed
G. E.	96	21½	Dismissed—reinstated	
L. S.	87	18	Dismissed	
	99	18	Dismissed	
E. E.	86	19	Dismissed	
Music	95	21½	Dismissed	
P. E. to Hist.	99	25	Dismissed	
Music	91	26	Dismissed	
G. E. to Hist.	77	28	Dismissed	

[1] Abbreviations: G. E., General Elementary; Kgn. P., Kindergarten Primary; L. S., Letter Sciences; H. E., Home Economics; Com., Commercial; P. E., Physical Education; Jr. H. S., Junior High School; St. Brd. Canad., State Board Credential Candidate.

The results of comparison of Alpha scores and grade point evaluations can be readily seen in the foregoing table:

More than eighty-two per cent of the group failed to maintain a "C average" during the two semesters and those who dropped out after the first mid-term examination ("mid-terms" come at six-week intervals) saw the warnings and took heed before the "faculty axe" fell. The results as brought out in the tabulation would tend to show that the students with Alpha scores of less than one hundred should be taken under advisement prior to matriculation.

A program embodying the following phases is advantageous:

1. Determination of causes for low scores, by retesting with Army Alpha, other group tests, or by individual examination. Of course in some cases the cause will be some unusual circumstance or handicap, *e.g.*, linguistic difficulties, or health at the time of testing.

2. Conference with student advising him of the chances of success and completion of collegiate work. This, of course, must take into account outside obligations (*e.g.*, family and self support), health, and maturity. A readjustment must often be faced by the student in terms of vocational interest and objective with the accompanying psychological and social phases.

In general, this study would tend to support the use of the Army Alpha in the classification of students for pre-registration advising, one of the most important phases of educational guidance.

NOTES ON ARTICLES IN EDUCATIONAL PSYCHOLOGY IN CURRENT ISSUES OF OTHER MAGAZINES

REPORTED BY JAMES E. MENDENHALL

Research Associate, Institute of Educational Research, Teachers College
Columbia University

Intelligence Testing

A Statistical Study of the Individual Tests in Ages Eight and Nine in the Stanford-Binet Scale. J. E. Wallace Wallin. Mental Measurement Monographs, June, 1929. The study included a testing of over 1000 school children.

Comparative Validity of Primary Intelligence Tests. Paul V. Sangren. The Journal of Applied Psychology, August, 1929, 392–412. Using the data from one hundred first grade children, the tests in order of ranking were: Haggerty Delta I, Pressey Primary Classification, Pintner-Cunningham, Detroit First Grade, Otis Primary, Kingsbury Primary, Rhode Island.

The Accuracy of the Abbreviated Stanford-Binet Intelligence Scale. Fowler D. Brooks. The Psychological Clinic, March–April, 1929, 17–20. Data from five hundred seventy-five individuals, ages six to seventeen, showed the abbreviated scale (1) to be from two to four points less accurate in IQ, (2) to have a lower average IQ, (3) to rank pupils relatively as on the complete scale, (4) to be less accurate with brighter individuals.

The Relation between Intelligence, Age and Home Environment of Elementary School Pupils. M. Sirkin. School and Society, August 31, 1929, 304–308. From 2733 pupils, Grades IV to VIII, intelligence test and social status ratings were secured. The correlation between (1) intelligence and social status was .40; (2) age and score, −.26.

Learning the Relation of Opposition. George Kreezer and Karl M. Dallenbach. The American Journal of Psychology, July, 1929, 432–441. One hundred children, kindergarten to second grade, were tested and practised in knowledge of opposites.

The Porteus Maze Test as a Discriminative Measure in Delinquency. Louise E. Poull and Ruth P. Montgomery. The Journal of Applied Psychology, April, 1929, 145–151. A comparison of seventy-two conduct disorder cases and eighty-one socially adjusted cases on the Porteus Maze showed the adjusted group to be slightly superior. On the Stanford-Binet and the Healy Pictorial Completion tests the two groups were equal in intelligence.

Mental Tests Findings with Armenian, Turkish, Greek and Bulgarian Subjects. Margaret M. Wood. The Journal of Applied Psychology, June, 1929, 266–273. In the Constantinople Woman's College, there were only slight differences in intelligence among Armenians, Turks, Greeks, Bulgarians.

An Investigation into the Validity of Norms with Special Reference to Urban and Rural Groups. Myra E. Shimberg. Archives of Psychology, Columbia University, March, 1929. The author tested large numbers of children from urban and rural communities with two information tests: Scale A, adapted to

urban children; Scale B, to rural children. Each group proved its superiority on the scale which included information adapted to its own environment.

Improvement of Feeble-minded Girls over Sixteen Years of Age. Margaret Otis. The Journal of Applied Psychology, June, 1929, 301–315. With training in language and book work, the scores on intelligence tests were increased.

Measurement of Achievement

Correlation Studies of the Student's Ability Index. Howard J. Banker. Journal of Educational Research, June, 1929, 31–37. The student's ability index (SAI) $\left(\frac{\text{grade mark}}{\text{chronological age}} \times \frac{\text{mental grade-age}}{\text{average grade mark}}\right)$. The SAI (Otis test) distributions and standard deviations compared favorably with those of IQ's by Otis and Binet (Terman); correlation of Otis SAI with Otis IQ's averaged .825. The SAI proved highly reliable, an average coefficient of .826.

Speech and Speech-reading Tests for the Deaf. Rudolf Pintner. The Journal of Applied Psychology, June, 1929, 220–225. From a study of four hundred cases, these results are secured: (1) Teacher's ratings of pupil's language is higher than that of other examiners; pupil's understand teacher's speech better than that of other examiner; (2) correlations between speech or speech-reading and Pinter Non-language score ranged from −.15 to +.16; (3) institutional cases exhibit lower mean speech, higher mean language, and greater variability than day school pupils; (4) pupils in institutions show lower mean and higher variability in speech reading than day school pupils; (5) correlations between Pintner Educational Survey Test and speech or speech-reading ranged from .32 to .65.

A Study of Diction Consciousness of College Freshmen. T. H. Schutte. Peabody Journal of Education, March, 1929, 289–293. Reports of a test differentiating knowledge of "correct" expressions.

New Standards in Written English. G. M. Wilson. The Elementary English Review, May, 1929, 117–119. Returns from a language testing program of 136,030 children, Grades III to XII.

Variety in Sentence-structures. Its Relation to Technical Excellence in Composition and to Intelligence. Irvin C. Poley. The Elementary English Review, May, 1929, 126–128. Variety of sentence structure (use of nouns, adverbial and adjective clauses, of introductory phrases, participial phrases, and inverted order) correlated with (1) IQ, .1; (2) mental test scores, .3; and (3) technique grades, .7 (*N*, 26).

The Validation of New Types of Plane Geometry Tests. Harry A. Greene and Ruth O. Lane. The Mathematics Teacher, May, 1929, 293–301. Data on standardization of the Lane-Greene Unit-achievement Tests in Plane Geometry. Reliability coefficients of .78 to .90 (*N*, 200).

Manuscript Writing Standards. Edith U. Conrad. Teachers College Record, Columbia University, April, 1929, 669–680. Description of the construction of a scale.

Educational Psychology

Changing Tendencies in General Psychology. Robert A. Davis and Silas E. Gould. Psychological Review, July, 1929, 320–331. The results of an analysis

of one hundred ten texts published between 1890 and 1928 showed tendencies in the field of general psychology from one decade to the next.

Left-handedness in Relation to Speech Defects. David W. Oates. The Forum of Education, June, 1929, 91–105. Seven tests were administered to 226 sinistrals and 3950 dextrals. The results indicated (1) higher per cent of speech defects among left than among right handed; (2) speech defects, six times as frequent in the mixed-handedness group as in the pure-handedness group; (3) nervous disability, two and one-half times as frequent in the "crossed" groups as in the "pure" groups.

A Study of Dextrality in Children. Julia Heil Heinlein. Journal of Genetic Psychology, March, 1929, 91–119. Children, ages four to twelve, were studied with reference to their preference in handedness, using dynamometer, steadiness, tapping and target tests. Comparisons between performances of right and left hand for successive ages are presented.

Practice and Improvement in the Müller-Lyer Illusion in Relation to Intelligence. H. R. Crosland, H. R. Taylor, and S. J. Newsom. The Journal of Genetic Psychology, April–July, 1929, 290–306. With thirty-three subjects, there was no significant correlation between size of error and intelligence, or in reduction of error through practice and intelligence.

The Maze-learning Ability of Blind Compared with Sighted Children. Josephene R. Knotts and W. R. Niles. Journal of Genetic Psychology, March, 1929, 21–50. With forty cases (paired), the blind proved superior and more variable in scores on the maze.

The Relative Potency of Color and Form Perception at Various Ages. Clara R. Brian and Florence L. Goodenough. Journal of Experimental Psychology, June, 1929, 197–213. A group of four hundred seventy-four subjects, ages fourteen months to adulthood were tested. Children below the age of three tend to match upon the basis of form rather than color; between three and six color in preference to form; at six and after form preference predominates increasingly.

An Investigation of Some Simple Speed Activities. Karl C. Garrison. The Journal of Applied Psychology, April, 1929, 167–172. Testing sixty college students, the writer reports correlations among these factors: mixed reading; continued reading; substitution of letters for numbers; cancellation of 4's, and of *c*'s, *i*'s *m*'s and *r*'s; marking tallies; card sorting and distribution; Otis SA; and form analogies.

An Experimental Study of the Perception of Objects. G. L. Freeman. Journal of Experimental Psychology, August, 1929, 341–358. Using sense and nonsense visual materials, the author finds that in addition to the emergence of figure from ground (Gestalt school), "a part (of the figure) becomes focal while the general shape serves as a background."

Initial Performance as a Basis for Predicting Ultimate Achievement. George W. Hartmann. School and Society, April 13, 1929, 495–496. Data obtained from fifty members of a class in educational psychology who practised the function of reciting the alphabet forward, intercepting n, backward and backward intercepting n.

Rate of Work in Schools. A Study in Variations. W. J. Stamer. British Journal of Psychology, April, 1929, 439–451. Using five hundred thirty-two pupils, the author determined the time during the day when the most work in

arithmetic computation was done; and also the period during the school term of maximum achievement.

An Experiment in Teaching Psychology. Norma U. Scheidermann. The Journal of Applied Psychology, April, 1929, 188–191. By use of a questionnaire the writer found that students preferred frequent short objective tests with knowledge of their status to long subjective tests at the end of the semester.

Objective Measures of Ability and Achievement in General Psychology. Howard P. Longstaff and James B. Porter. The Journal of Applied Psychology, April, 1929, 173–187. Tests of two hundred seven university students showed these correlations: (1) Average grade in fourteen objective tests in general psychology and (*a*) May's "A Standardized Eaxmination in General Psychology," .491; (*b*) Otis S. A., .91; (*c*) average grade in college work excluding psychology, .584; (*d*) average grade in high school course, .280. (2) Correlation between May's test and (*a*) Otis S. A., .338; (*b*) college grades, .435; (*c*) high school grades, .333. (3) Correlation between Otis S. A. and (*a*) college grades, .334; (*b*) high school grades, .148. Correlation between college grades and high school grades, .423. Partial and multiple correlations and regression equations are presented. "Objective tests as used in this study agree better with college averages than do either high school averages or intelligence test scores."

A Study in Phonetic Symbolism. Edward Sapir. Journal of Experimental Psychology, June, 1929, 225–239. An interesting experiment to determine the significance to children and adults of long and short sounds apart from the conventions of acquired language and meaning.

A Study of the Development of the Rate of Articulation. L. O. Dawson. The Elementary School Journal, April, 1929, 610–615. The article includes norms for sexes and grades I to XII on the following tests: Number reading, alphabet, word, repetition of nursery rhyme, repetition of an alliterative sentence and a sentence test. Also characteristic scores of individuals atypical in intelligence.

The Bearing of Professor Pavlov's Work on the Problem of Inhibition. William McDougall. The Journal of Genetic Psychology, April-July, 1929, 231–262. This review of Pavlov's recent book stresses the concept of "inhibition by drainage.

Is Thorndikianism the Last Word in the Psychology of Learning? Hsiao Hung Hsiao. The Journal of Genetic Psychology, January, 1929, 115–128. A critical essay.

The Net Result of the Anti-heredity Movement in Psychology. Zing Yang Kuo. The Psychological Review, May 1929, 181–199.

Psychology of Learning and School Subjects

An Experiment in the Use of Two Methods of Instruction. Martha Willard. Educational Method, June, 1929, 505–510. Learning of history by four hundred eleven sixth grade children who were divided into two equal groups. The experimental group used a modified Dalton plan of study; the control, the daily recitation procedure. The former proved superior.

A Practicable Technique for Determining the Relative Effectiveness of Different Methods of Teaching. Ernest O. Melby and Agnes Lien. Journal of Educational Research, April, 1929, 255–264. The experimental group (*N*, 25) which was taught by the project method was inferior in gains to each of the control groups (*N*, 87) on a fact-knowledge test in geography.

Achievement of Dull Pupils under the Dalton Plan. Harold H. Abelson. School and Society, August 10, 1929, 211–212. With one hundred sixty-nine pupils (Grade VI) intelligence tests were used to form two groups of equal ability. The superiority of the Dalton plan in all classes was indicated.

The Correlation of Laboratory and Classroom Work in the Teaching of High School Chemistry. Grace Bagby. Journal of Educational Research, May, 1929, 336–340. Three groups (*N* each, 30) were taught for thirteen weeks by each of three methods: (1) No class work before laboratory; (2) one day preview in the class before laboratory; and (3) completion of the topic in the classroom followed by experiment. No significant differences were found. For pupils of low or median intelligence—"utilize the first day for class discussion, the second, for experiment, the third for class work." For those of high intelligence, complete "the subject-matter in the class before going into the laboratory or by doing the experiment before having any classwork whatever."

Individual Instruction in Vocational Agriculture. Roy W. Roberts. Journal of Educational Research, May, 1929, 349–352. The group in industrial instruction (use of job sheets, etc.) proved decidedly superior to that in group instruction.

Experimental Comparisons of Print-script and Cursive Writing. Arthur I. Gates and Helen Brown. Journal of Educational Research, June, 1929, 1–14. (1) In the first grade (*N*, 44) print-script proved superior to cursive writing in ease of learning (speed and quality). (2) In Grades II to III, print-script (which had been taught in school previous to experiment) was superior in quality and decidedly superior in speed to cursive writing. From Grades IV to VI print-script (after period of drill) was slightly superior in quality but inferior in speed to cursive writing. Two classes in the sixth grade were practised in print-script and cursive writing; both gained equally in quality, but the cursive group gained decidedly more in speed. Print-script is preferred but cursive is increasingly "liked" from Grades II to VI. Children (Grades I to III) who had learned print-script learned cursive readily during a short period of practice.

Factors Affecting Ability in Handwriting. Milton Wittler. School and Society, June 29, 1929, 847–850. A study of four hundred ninety cases, ages ten to eleven, to determine factors associated with good handwriting revealed the following correlations; (1) Achievement Index with rate, .21; with quality, .18; (2) IQ with rate, .35; with quality, .31; (3) girls were slightly superior to boys in AI, IQ, in rate and quality of writing.

Sex Differences in Handwriting. M. Eustace Broom, Blanche Thompson, and M. Thelma Bouton. The Journal of Applied Psychology, April, 1929, 159–166. "It is possible to determine sex of writers correctly roughly two times in every three judgments when the sex of the writer is unknown."

The Influence of Past Experiences upon Spelling. Ernest Horn. Journal of Educational Research, April, 1929, 283–288. In attempting to write the word "circus," one hundred ninety-five pupils in Grade I and II misspelled the word in one hundred forty-eight different ways; "tease" in forty-four ways. Likewise, "miscellaneous" was mispelled in one hundred thirteen different ways by eighth grade pupils; in forty ways by twelfth grade students. The author concludes that there is no hard spot of a word that is common to many pupils; therefore, the learning of each word is an individual problem.

The Reading Span. W. H. Pyle. The Elementary School Journal, April, 1929, 597–602. A tachistoscope experiment (3431 children, ages six to fourteen) showed sex, age and grade differences in reading span.

Summary of Arithmetic Investigations (1928). Guy Thomas Buswell. The Elementary School Journal, (I) May, 1929, (II) June, 1929, 737–747.

The Specific Type of Problem in Arithmetic versus the General Type of Problem. Claude Mitchell. The Elementary School Journal, April, 1929, 594–596. An average difference in favor of specific problems was 2.5; the correlation between specific score and general score, .52.

Specific versus General Drill in the Fundamentals of Arithmetic in Grade IV. L. D. Morgan. School Science and Mathematics, May, 1929, 528–529. Diagnostic and remedial test exercises prove superior to practise in general material.

An Analysis of the Learning Units in N Processes in Algebra. The Mathematics Teacher, May, 1929, 245–283. Comprehensive tests of three hundred fifty first year algebra pupils are analyzed for difficulties. Extensive tables of errors are presented.

Studies in the Learning of English Expressions No. II: Capitalization. Percival M. Symonds and Baldwin Lee. Teachers College Record, April, 1929, 686–692. A report of an extensive analysis of errors in capitalization.

Curriculum

Leisure Reading of Junior High School Boys and Girls. Joe Jennings. Peabody Journal of Education, May, 1929, 333–347. A study of the reading done by eight hundred ninety children during a week's time. The newspaper reading, magazines and books were classified by type separately for boys and for girls.

Reading Interests of High-school Children. Vera Elder and Helen S. Carpenter. Journal of Educational Research, April, 1929, 276–282. This study of 4000 girls enumerated the books, magazines, and reasons for liking particular books.

The Reading Interests of Adults (Cleveland). Alonzo G. Grace. Journal of Educational Research, April, 1929, 265–275. An analysis of the personal record cards of 3289 adults showed interesting facts regarding remote ancestry, subjects and types of reading; books, magazines, and newspapers read.

The Three-year Curricula in Normal Schools for the Preparation of Kindergarten-primary Teachers. Charles H. Thompson. Journal of Educational Research, May, 1929, 322–330. The three-year course in fifteen schools proved to be merely an extension of the two-year course which was in need of some modification.

Informational Mathematics versus Computational Mathematics. Charles H. Judd. The Mathematics Teacher, April, 1929, 187–196.

Materials of A Historical Nature Contained in Twenty-six Primary Readers. Mildred A. Dawson. The Elementary School Journal, June, 1929, 756–766. "The amount of space in twenty-six readers which is devoted to historical material is small, but it tends to increase from grade to grade."

A Study of the Vocabulary of Ten Spellers. Erich Selke. The Elementary School Journal, June, 1929, 767–770. "Two-thirds of the words common to all ten spellers are placed in two or three successive grades."

Conclusion of the Vocabulary Study of English and Metric Units of Measure. Oscar Hunebaugh. Educational Research Bulletin, April, 17, 1929, 175–180. Four criteria to determine units needed were employed: Science textbooks, maga-

zines, home check list, and arithmetics. These units were submitted to high school pupils in the form of a test.

Administration and Supervision

A Comparison of Normal School Students with Certain Other Groups. Verner Martin Sims. School and Society, July 13, 1929, 68–70. In comparison with college freshmen, normal freshmen (*N*, 67) were (1) inferior on Carnegie Tests for General Mental Ability, Cross English, Thorndike Vocabulary, and Sims Score Card (twenty to twenty-six per cent of normal group reaching median of college group). In comparison with demonstration high school, adult norms, and ninth grade norms for Stanford Achievement, the normal freshman equalled the ninth grade norms.

Marking Systems in Teacher-training Institutions. Homer L. Arnold. California Quarterly of Secondary Education, April, 1929, 244–248. Returns from one hundred institutions in thirty-four states were utilized. Wide variation in practice was found.

Teachers College Students' Estimates of Their Own Scholastic Merit. School and Society, August, 24, 1929, 275–276. There was a fairly close agreement between student's estimate and instructor's estimate of scholarship (*N*, 34).

The Effect of Absences on Scholarship. Blake Crider. School and Society, July 6, 1929, 27–28. The records of two hundred forty college freshmen disclose this fact: "A student's scholarship is affected both by his percentile ranking on the psychological test and (but much less) by his absences."

Class-size Conditions and Trends at the College Level. Earl Hudelson. School and Society, July, 20, 1929, 98–102. A report from three hundred eighty-six institutions indicated that non-degree-granting teachers colleges and junior colleges have the largest classes; state colleges and state universities, the smallest classes. The largest classes were in required courses in social sciences and English; the smallest in highly specialized courses, in electives and in mathematics.

Some Aspects of Current Efforts to Improve College Instruction. Floyd W. Reeves and John Dole Russell. Bulletin of the Bureau of School Service, University of Kentucky, December, 1928. A summary of recent studies.

Students' Opinions of Their Best and Poorest Teachers. David Newmark. The Elementary School Journal, April, 1929, 576–585. Data from two hundred twenty-three normal school students were used to determine traits of best teachers (high school) and those of poorest teachers. Skillful in "getting ideas across" was the trait rated most highly.

Departmental Differences in the Quality of Instruction as Seen by Students. H. H. Remmers. School and Society, September 7, 1929, 332–334. The study shows great variability within a department, and but little variation among departments, in the students rating of traits of professors.

Results of the Elective System in Denver. A. K. Loomis. The School Review, September, 1929, 510–518. A study of the records of 2904 high school graduates showed the subjects and units in each taken, correlations of subjects and units taken, etc.

A Technique for Analyzing the Supply and Demand of Educational Workers. Albert Grant and W. H. Cowley. School and Society, May 11, 1929, 618–620

Tables showing supply and demand for teachers in various types of work, in subjects, and in related fields.

Office Records and Reports in Secondary Schools. W. C. Reavis and Robert Woellner. The School Review, September, 1929, 496–509. An analysis of the records from five hundred twenty-two representative schools to determine common items on pupils' record cards, teachers' records, records of equipment and supplies and reports to boards of education, etc.

Salary Scales in City School Systems. 1928–1929. Research Bulletin of the National Education Asssociation, Washington, D. C., May, 1929.

Tenure in Present Positions and Training of City School Superintendents. Margaret M. Alltucker. School and Society, May 18, 1929, 649–652. An analysis of the records of 1679 superintendents showed: (1) "The median tenure of superintendents of city schools in their present positions is seven years," and (2) "Of fifty-eight superintendents of schools in cities over 100,000 in population, twenty-three or forty per cent, have doctors' degrees."

The Principal and Progressive Movements in Education. Research Bulletin of the National Education Association, Washington, D. C., March, 1929.

The Training of Teachers of the Modern Foreign Languages. C. M. Purin. The Macmillan Co., New York, 1929.

Character and Personality Traits

An Objective Measure of Emotional Instability. Robert Jaudon Ball. The Journal of Applied Psychology, June, 1929, 226–256. Time and errors for tracing a maze were used as a measure of emotionality.

Experimental Studies in Suggestion. G. H. Estabrooks. Journal of Genetic Psychology, March, 1929, 120–139. The tests (electricity, bob, box, a group test of observation, and psychogalvanic reflex) correlated little with one another or with age and intelligence.

Electrical and Circulatory Responses to Brief Sensory and Ideational Stimuli. Chester W. Darrow. Journal of Experimental Psychology, August, 1929, 267–300. Change in blood pressure as responses to sensory, disturbing ideational and indifferent ideational stimuli.

The Emotional Behavior of Young Children during Mental Tests. Florence L. Goodenough. The Journal of Juvenile Research, July, 1929, 204–219. A rating scale. Age differences, sex differences, and intercorrelations for trait ratings are included in the article. Nine hundred ninety children, ages eighteen months to five years eleven months were rated in *shyness*, *negativism* and *distractibility*.

The Measurement of Conative Factors in Children and Their Influence. C. A. Richardson. The British Journal of Psychology, April, 1929, 405–412. The cases (two hundred forty-four children, ages eleven to thirteen) were tested with the Downey-Will-Temperance Test, and intelligence test and tests of attainment (to the eleven-year group only).

Emotional Factors in Learning. Harold Ellis Jones. The Journal of Genetic Psychology, April-July, 1929, 263–272. From Jung's list, sixteen words were selected. These were administered to seventy-six women and thirty-eight men, psychogalvanic responses being recorded. Words with high P. G. R. effect such as "kiss, love, woman" were better remembered than those with low effect, "afraid, flower, glass."

Periodic Emotional Changes in Male Workers. Rex B. Hershey. The Personnel Journal, April, 1929, 459–464. Workers' ratings of their feelings (scale of thirteen points) toward work from day to day show evidence of periodicity.

Optimism and Pessimism in College Environments. Herbert H. Jasper. The American Journal of Sociology, 856–873. A questionnaire filled out by five hundred fifty-four university and college students to determine the attitudes toward personal and social conditions both in and outside of college. The writer ranked schools in order from optimistic to pessimistic attitudes thus—normal school, a denominational school, state university, private college.

A Critical Study of a Test of Extroversion-introversion Traits. The Journal of Juvenile Research, April, 1929, 104–123. A study of sixty-two men and three hundred twenty-nine women showed a fairly high validity and a low reliability for the test.

Errors in Self Judgment. T. A. Jackson. The Journal of Applied Psychology, August, 1929, 372–377. The correlation between intelligence (concensus) and "ability to judge oneself" was .70; and between intelligence (by test) and this ability was .48. In comparison with five least intelligent, the five most intelligent displace themselves (1) decidedly less in all eight traits; (2) slightly more in undesirable traits, etc.

The Reliability of Self-scored Measures. Lloyd N. Yepsen. School and Society, May 25, 1929, 681–682. With fifty-nine cases, the study showed that students below the median of achievement were more dishonest than those above the median.

The Effect of Character Traits on Scholastic Achievement. H. J. Steere. School and Society, June 1, 1929, 707–709. Comparisons between groups of "high" character rating and groups of "low" rating in intelligence and achievement showed that desirable traits are positively associated.

Ideas on Right and Wrong among Teachers and Children. Vernon Jones. Teachers College Record, March, 1929, 529–541. Report of responses to a "test of ethical choices" administered to one hundred seventy-seven children in the seventh and eighth grades and to one hundred fifty-nine adults showed lack of agreement by both adults and children as to the "ethical" course of action.

An Improved Method of Detecting Cheating in Objective Examinations. Charles Bird. Journal of Educational Research, May, 1929, 341–348. Three methods for determining cheating by comparing errors of neighbors were used. A short method consisted of comparing the like errors of the suspect with (1) his neighbor and (2) with ten other students of same scholarship rating on the test.

The Interpretation of Facial Expression in Emotion. Carney Landis. The Journal of Genetic Psychology, January, 1929, 59–72. Using photographs of facial expressions with forty-two judges, "The names given by the observers to the expressions were no more often correct than one would expect by chance."

Statistical Methods

The Predictive Value of Mental Tests that Satisfy Spearman's Tetrad Criterion. E. J. Asher. The Journal of Applied Psychology, April, 1929, 152–158. Tables of correlations, tetrad differences, and validity coefficients are presented.

The Interpretation of the Correlation Coefficient. Robert Choate Tryon. Psychological Review, July, 1929, 419–445. "The methods will be given whereby

by knowing rxy one may discover the per cent (or the limits of this per cent) by which Y is determined by X_1 and also the remaining per cent Y is determined by residual factors other than X which make up the rest of Y."

Fechner's Law and the Method of Equal-appearing Intervals. L. L. Thurstone. Journal of Experimental Psychology, June, 1929, 214–224. The experiment (N, 101) involved the sorting of ninety-six cards each of varying number and placement of dots, into ten piles of equally apparent density. Using Fechner's equation (S equals K log R), the statistical data were compared with the experimental data.

Theory of Attitude Measurement. L. L. Thurstone. The Psychological Review, May, 1929, 222–241.

The Sampling Problem in Educational Research. Helen M. Walker. Teachers College Record, May, 1929, 760–774. Examples from fields of study are cited as violations of the rules of simple sampling.

Statistics in Social Research. Dorothy Swaine Thomas. The American Journal of Sociology, July, 1929, 1–17. The possibilities and limitations are discussed.

NEW PUBLICATIONS IN EDUCATIONAL PSYCHOLOGY AND RELATED FIELDS OF EDUCATION

CONDUCTED BY FRANCES M. FOSTER

Scientific Method: Its Function in Research and Education, by T. L. Kelley. Columbus: Ohio State University Press, 1929. Pp. VII + 195.

Philosophical mathematicians like Poincaré, Pearson, Russell and Whitehead seem prone to the production of treatises important from the standpoint of education. To this list Kelley's name must now be added, for the work under review is probably more important for advanced students of education, particularly those beginning research for a Ph.D. degree, than Pearson's "Grammer of Science" or Poincaré's "Science and Hypothesis." Consisting only of five lectures delivered at Ohio State University, Kelley's "Scientific Method" will have abiding value. The titles of the lectures are: (1) Interrelationship between Method of Research and Field of Investigation; (2) The Rôle of Judgment in "Objective Measurement"; (3) In What Units Shall We measure Intelligence and Achievement? (4) The Bearing of Recent Scientific Development upon Problems of Education and Inheritance; and (5) Mental Traits of Men of Science.

The treatment is surprisingly fresh. Throughout the book new material is continually added, and even when the material is old some novel treatment is sure to make it interesting. His analysis of historical method, experimental method, and the rôle of judgment is clear and succinct. The questionnaire study of the second lecture has intrinsic merit, besides enabling the lecturer to show the scientific value of a consensus of opinion. The fourth lecture while starting with curriculum construction and ending with eugenics nevertheless hangs well together.

Selecting Chap. III for a more extended discussion because of its more immediate bearing upon practical problems of the day, we note that he discusses the meaning of nature and nurture. Out of this discussion he develops methods for scaling tests more accurately, and demonstrates these upon findings from the Stanford Achieve-

ment Tests and the Stanford Binet. His methods are those of sensed differences (by grade teachers), an ingenious application of Weber's Law; and that of maximum correlation of an individual's score at different ages, a development of his theory of nature and nurture. The results from the two methods are in striking agreement. If the grade differences between III and IV are 1 then those between VII and VIII give values of .43; .44, .45, .50, .66 and .83; while those between II and III give values of 1.66 and 2.0. This agreement he (inferentially) regards as supporting his theories of nature and nurture as independent variables. They can, however, be made to support another one. If nature be regarded as the sum total of the traits (or genes) in the fertilized ovum, and nurture the factor determining which of these potentialities shall be made actual, then a uniform environment (posited by intelligence and achievement testers) will produce an average growth curve, measured in absolute units, that is logarithmic in character. Thorndike, for example, has shown this to be true for intelligence. The intervals for Grades II and III will be greater than those for III and IV and these again greater than those for VII and VIII. This is what Kelley finds, the variations being due to the imperfect nature of the sampling both of test items and of subjects. This view of the correlative nature of heredity and environment, this 50-50 view of their contributions, is simpler than Kelly has propounded and therefore, if it explains the facts, should be adopted. Anyhow, it is certain that the distinction between an intelligence test and an achievement test is artificial. Both are achievement tests, although we use the results of an intelligence test as a measure of potentiality because of the assumption that a uniform environment has developed different heredities to different extents.

The reasoning employed by Kelley is very keen and needs close attention. The book cannot be skimmed like a novel; its material is intrinsically difficult and every word needs to be read carefully. This does not mean that the book is dull—far from it. To meet a statement like the following: "Generally speaking, the practice of the nations of the western world is to reward specialists; and their philosophy, as reflected in their public school curricula, is to develop paragons of uniformity" is to remember it forever.

The proof-reading seems to have been hurriedly done. When other editions are called for (and they certainly will be, since every wise director of Ph.D. students will require his students to read the book) the mistakes should be corrected. The more obvious ones are

the sentence ending the second paragraph on p. 10: seismologists, instead of meteorologists, p. 35; the ambiguous sentence beginning "Lamarck and Weismann," p. 158; and the statement that Galileo invented the telescope, p. 163.

Kelly deserves the thanks of us all. The reviewer extends to him congratulations on the production of a book that has real significance.

PETER SANDIFORD.
University of Toronto.

The Nature and Direction of Learning, by W. H. Burton. New York: D. Appleton and Co., 1929. Pp. XVIII + 595.

This is a difficult book to review. In many respects it is of the class that Americans produce better than other people and of which they may be justly proud. It is workmanlike. From start to finish it shows evidence of care in preparation. The exploratory and pretest questions are excellent, and, in the very best sense, thought-provoking. The discussion questions at the end of each chapter, if worked faithfully by the student, will ensure him a thorough grasp of the material dealt with by the author. The illustrative examples are both apposite and humorous. Since it has been in preparation over a number of years and during the last three has been the subject of a critical and experimental attack at the University of Chicago, all the corners have been rubbed smooth. Moreover, with the exception of "Learning as Conditioning" all the aspects of learning—methods, principles, philosophies and psychologies—from Herbart's "Five Formal Steps" to Kilpatrick's "Project Method" and Morrison's "Unit Learning" here find a home. The wealth of the material is almost overwhelming.

Yet having made all the above favorable comments, the reviewer feels vaguely dissatisfied with it. In spite of a good deal of space devoted to them, the reviewer opines that the author of the book does not believe in Morrison's units or the permanency of his "adaptations." It is this intellectual insincerity permeating the whole of the book that strikes the wrong note with the reviewer. The criticism is a personal one and may be unjust but it must be made.

If the author from his wide studies, could have evolved a satisfactory theory of learning and surely there is a golden unifying thread to be found somewhere (conditioning perhaps), or even if he had

given the results of experiments in learning (which he does not), the book would have been an outstanding contribution to the subject. Even in its present stage of development it is certain to prove valuable and to be widely used.

PETER SANDIFORD.
University of Toronto.

Difficulties in Child Development, by Mary Chadwick. New York: The John Day Co., 1928. Pp. 411.

Social Problems of Childhood, by Paul Hanly Furfey. New York: Macmillan Co., 1929. Pp. X + 228.

The problem of thwarted childhood has its individual consequences as well as social implications and hence, can be written about from the point of view of two distinct levels of description. One is largely psychological in nature and represents an attempt to apply knowledge and techniques relevant for dealing with problems of child guidance. The other level of description is largely sociological in nature and the interest is in problems of child welfare as an aspect of public welfare. The book by Miss Chadwick is concerned with the first level; the book by Dr. Furfey is concerned with the second. Both books are crammed with interesting data. Neither is characterized by a disinterested presentation. An extreme gullibility of psychoanalytic thinking characterizes one; a predetermined faith serves as dynamic background for interpretation in the other.

"Difficulties in Child Development" is an attempt to condense the facts of psychoanalysis which contribute "towards the understanding and early education of young children" and popularize them for use of parents, teachers, nurses, and welfare workers. A bibliography appears in the front part of the book and a small number of references accompany each section. The book is divided into twelve sections. Characteristic captions are: The Child Discovers the World Around, The Great God-wish, Forgetting and Remembering, Dawn of Personality, Children's Games and Phantasies. The content of this book impresses the reviewer as a jumble of scientific and metaphysical physiology, of relevant and irrelevant anthropology, of reliable and highly unreliable psychology; a confusion of well ascertained facts and loose generalities. Characteristic of her style of writing and her cocksure manner of expressing facts or opinions is the following quotation: "The child, who, because of too much

fear of losing love or through too much pain in the process, is unusually quickly 'broken in' to habits of cleanliness, will be the suggestible, easily led, and readily influenced child of the future, always ready to adopt every suggestion made by another for fear of displeasing that person when present, or ready to surrender to the lightest bodily stimulus in order to relieve its feelings in any way." Her ready acceptance of recapitulation theory for explaining the date of appearance of native tendencies, her conclusion to the effect that the genesis of optimism is in breast feeding, her easy going statements about the nature of the only child, the hostility of the sexes, the patience of urethral children are among the many statements found in the book which are either known to be false or else very little susceptible to either experimentation or verification.

In "Social Problems of Childhood" the various aspects of child welfare are discussed in separate chapters. The first chapter is a general discussion of the child and the community and the last is an interpretative description of child-caring institutions. Child health, mental hygiene, illegitimacy, subnormality, dependency and child labor are other aspects considered. Each chapter is followed by a well-selected bibliography and each book included is briefly and intelligently commented on. The book has an index. The writing is clear and an analysis of the content indicates a familiarity with best practices in each of the applied aspects treated and a knowledge of social research as well as an appreciation of the need for it. The facts, practices and attitudes advocated by social scientists, however, the author eagerly seeks to prove in conformity with Catholicism. Because, as he tells us on the first page of the preface he "cannot too strongly stress the conviction—shared by all Catholics—that the neglect of the teachings of Christ is a supremely important factor underlying all social problems," he sees the problems as spiritual in nature and though he informs us that he is reserving the treatment of the churches' activities to a future work and in this book treats only of the community's share in the problems of childhood, the religious point of view colors his present work at its inception—the statement of the problem—as well as the consideration of it. It is doubtful if many well informed sociologists will agree with the author that "the neglect of religion is the most fundamental cause of crime." Nor will they agree with his attitude towards the values of volunteer work, nor his acceptance of an oversimplified conception of psychoanalysis. A good work but in his enthusiasm for things religious the

author forgets that, as Lippmann tells us in his "Preface to Morals," "the essence of scientific method is a determination to investigate phenomena without conceding anything to naive human prejudices. Therefore, genuine men of science shrink from the attempts of poets, prophets, and popular lecturers to translate the current scientific theory into the broad and passionate dogmas of popular faith."

H. MELTZER.

Psychiatric Clinic, St. Louis.

The Education of Mentally Defective Children, by Alice Descoeudres. Translated from the Second French Edition by Ernest F. Row. Boston, Mass.: D. C. Heath and Co., Pp. 312. $2.00.

In Europe, as in America, the educational problems of the mentally defective child have stimulated serious thought and intensive research. The study of the defective child from an educational point of view has long engaged the attention of Mlle. Descoeudres, who, in this second edition of a book published a number of years ago, presents her view-point at to aims and methods in the training of mentally defective children. Her description of methods is prefaced with a brief history of the pedagogy of subnormal children, a discussion of examination methods, medical, psychological, and pedagogical, and the details of school and class organization. An excellent outline is given for obtaining information about the child's personal characteristics and his home background. The author advocates a three-fold program for the training of special class teachers: scientific training for the study of the children to be instructed, technical training in special phases of special class instruction, and practical training in conducting such classes.

In the education of mental defectives the author believes that instructors must "contrive to combine a program of didactic instruction intended to provide the pupil with the knowledge that is indispensable for life with a program of psychological gymnastics, which aims at exercising the various mental functions and their reciprocal actions," and that the pupils should not be taught" ideas however interesting, but lessons in attention, in volition, and in discipline." She advocates natural activity for the child, the necessity of his using his body, hands and brain. There is need, she believes, for maximum freedom compatible with good discipline. Much impor-

tance is attached to perceptual knowledge and sense training, and the need for correlating subject-matter is emphasized. Handwork is valuable, in the author's opinion, for the training of attention, interest, will-power, honesty, self-confidence. Exercises in folding, cutting, pasting, she considers worth while for training these capacities. She advocates object lessons for sense training, speech development, and opportunity for work in arithmetic, drawing, handwork, physical training. Much space is given to speech training, its psychology and pedagogy, and the study of speech defects. For the correlation of speech defects the author advocates breathing exercises, and practice in phonation and articulation. Reading is taught in the beginning stages through the training of visual and auditory perception by means of "lotto" games. The aims of instruction are the strengthening of perception, imagery, association. A warning is given against beginning reading before the child has the necessary mental development for it. The Decroly number games are advocated for training the child in the acquisition of number concepts, chiefly through training sensation and imagery. For the moral training of the child the author would depend largely upon the teacher's example, the child's interest in his work, and reliance on the child's ability to carry responsibilities.

The reader is not always sure that Mlle. Descoeudres practices what she preaches. Although she advocates freedom for the child and opportunity for natural expression, much of her program, when expanded, impresses one as artificial and obsolete. The author's methods and interpretations are not in harmony with methods of of curriculum organization and instruction advocated by modern educators. Some of Mlle. Descoeudres' statements give us the impression that faculty psychology still prevails in education, that transfer of training is still depended upon to make artificial material and exercises educationally valuable, that the child is the passive receptor of the teacher's efforts, that subject-matter organization need have little relation to the child's natural activities. The author is almost apologetic about the inclusion in the curriculum of games more strictly utilitarian than those designed for sense training and perception.

The author's account of her methods probably does not convey an adequate picture of the real spirit of the classroom and the provisions made to insure each child's best individual development. The

joyous experiences of the children on their excursions outside the classroom are happily portrayed. GERTRUDE HILDRETH.

The Lincoln School of Teachers College.

The Questionnaire in Education. Leonard V. Koos. New York: Macmillan Co., 1928. Pp. VII + 178.

The last few years has seen the publication of several books on methods of research and experimentation in education. Fairly adequate sources are available as guides to methods of experimentation, measurement, use of the case method and statistics. Koos', "The Questionnaire in Education," is an important manual to be added to this list.

Professor Koos has achieved a two-fold contribution in the book under consideration. First, he has shown, through a detailed analysis of five hundred eighty-one studies in recent periodicals and monographs, the purposes, extent, and values of the use of the questionnaire as a source of scientific data. Having done this, the analysis yields an excellent basis for generalization in regard to adequate methods of questionnaire construction.

The importance of the questionnaire as a source of data is evident when it is seen that practically one-fourth (24.6 per cent) of the studies examined utilized this method. It is on a par with the other two most prevalent methods of collecting data, namely, documentary analysis (25.0 per cent) and experimentation (23.6 per cent). These proportions were approximately the same when the analysis was made separately for the periodical literature, on the one hand, and for the research studies embodied in the "Teachers College Contributions" and the "Supplementary Educational Monographs," on the other hand.

The first three chapters bring out the above facts. The remaining chapters, IV and VII inclusive, are concerned with the details regarding the type of response, criteria of questions and questionnaires, means of stimulating replies, validation, resumé and recommendations.

The book is interestingly written. The discussion follows a well organized outline and the tables and graphs are very clearly printed and arranged. A number of illustrations of faulty types of questions is given. If it had been possible to include a collection of some of the best questionnaires as illustrations it would have added to the

manual. However, a bibliography is appended directing one to such examples, as well as to articles dealing with the questionnaires method.

In the light of the author's aim a significant contribution has been made. "The aim has been to prepare a book of compact proportions that will be of interest and value to students of educational research and to that large body of educational workers who undertake questionnaire studies."

C. O. MATHEWS.
Ohio Wesleyan University.

CORRECTIONS

The article, "Environmental and Character Factors Involved in Scholastic Success: 1926–1927 Data," by Frank K. Shuttleworth, in September Journal contains an error in the order of paragraphs. On page 431, the paragraph beginning "It seems worth while" should precede the paragraph at the top of page 430 beginning "Again, the intercorrelations."

In the article, "The Measurement of Achievement in General Inorganic Chemistry" by Victor H. Noll, in the September issue, the following corrections should be made:

Page 463, line 6, $k = .340$ instead of .436.

Line 8, page 463, should read, "six hundred and sixty" instead of "five hundred sixty-four."

Also Table III, No. 12—$r = .932$, PE of $r = .013$.

Likewise Page 464, paragraph 4, line 6, and page 465, paragraph 2, line 3, the correlation is .932 instead of .885.